UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

OFFICE OF
CHEMICAL SAFETY AND POLLUTION
PREVENTION

MAR 16 2017

MEMORANDUM

SUBJECT: Transmission of Meeting Minutes and Final Report of the December 13-16, 2016 FIFRA SAP Meeting Held to Consider and Review Scientific Issues Associated with EPA's Evaluation of the Carcinogenic Potential of Glyphosate

TO: Rick P. Keigwin, Jr.
Acting Director
Office Pesticides Programs

FROM: Steven M. Knott, M.S.
Acting Executive Secretary
FIFRA SAP Staff
Office of Science Coordination and Policy

THRU: Stanley Barone, Ph.D.
Acting Director
Office of Science Coordination and Policy

Please find attached the meeting minutes and final report of the December 13-16, 2016 Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) open public meeting held in Arlington, Virginia. This report addresses a set of scientific issues associated with EPA's evaluation of the carcinogenic potential of glyphosate.

Attachment

Plaintiff Exhibit
0762

cc:

Wendy Cleland-Hamnett
Louise Wise
Stan Barone
Arnold Layne
Delores Barber
Marietta Echeverria
Michael Goodis
Yu-Ting Guilaran
Steve Knizner
Robert McNally
Wynne Miller
Jaqueline Mosby
Dana Vogel
Gregory Akerman
Jeff Dawson
Anwar Dunbar
Anna Lowit
Cathy Milbourn
Monique Perron
Linda Strauss
OPP Docket

FIFRA Scientific Advisory Panel Members

Marion F. Ehrich, PhD, DABT, ATS
David A. Jett, PhD
James McManaman, PhD
Joseph Shaw, PhD
Sonya K. Sobrian, PhD

FQPA Science Review Board Members

Kenny Crump, PhD
Laura C. Green, PhD, DABT
Eric S. Johnson, MB; BS (MD), PhD, MPH, DTPH
Barbara L. Parsons, PhD
Kenneth Portier, PhD
Aramandla Ramesh, PhD
Elizabeth A. (Lianne) Sheppard, PhD
Emanuela Taioli, MD, PhD
Daniel Zelterman, PhD
Luoping Zhang, PhD

FIFRA Scientific Advisory Panel
Meeting Minutes and Final Report
No. 2017-01

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

EPA's Evaluation of the Carcinogenic Potential of Glyphosate

December 13-16, 2016
FIFRA Scientific Advisory Panel Meeting
Held at the EPA Conference Center,
One Potomac Yard
Arlington, Virginia

TABLE OF CONTENTS

NOTICE

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) is a Federal advisory committee operating in accordance with the Federal Advisory Committee Act and established under the provisions of FIFRA as amended by the Food Quality Protection Act (FQPA) of 1996. The FIFRA SAP provides advice, information, and recommendations to the Agency Administrator on pesticides and pesticide-related issues regarding the impact of regulatory actions on health and the environment. The Panel serves as the primary scientific peer review mechanism of the Environmental Protection Agency (EPA), Office of Pesticide Programs (OPP) and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. Food Quality Protection Act (FQPA) Science Review Board members serve the FIFRA SAP on an *ad hoc* basis to assist in reviews conducted by the Panel. These meeting minutes and final report have been written as part of the activities of the FIFRA SAP and represent the views and recommendations of the FIFRA SAP and do not necessarily represent the views and policies of the EPA, or of other agencies in the Executive Branch of the Federal government. Mention of trade names or commercial products does not constitute an endorsement or recommendation for use. The meeting minutes and final report do not create or confer legal rights or impose any legally binding requirements on the EPA or any party. In preparing the meeting minutes and final report, the FIFRA SAP carefully considered all information provided and presented by the EPA, as well as information presented in public comments.

These meeting minutes and final report of the December 13-16, 2016 FIFRA SAP meeting held to consider and review scientific issues associated with EPA's evaluation of the carcinogenic potential of glyphosate were certified by James McManaman, Ph.D., FIFRA SAP Chair and Steven Knott, M.S., Designated Federal Official. The minutes and final report are publicly available on the SAP website (https://www.epa.gov/sap) under the heading of "Scientific Advisory Panel Meetings" and in the public e-docket, Docket Identification Number: EPA-HQ-OPP-2016-0385, accessible through the docket portal: https://www.regulations.gov. Further information about FIFRA SAP reports and activities can be obtained from its website at https://www.epa.gov/sap. Interested persons are invited to contact Steven Knott, Designated Federal Official, via email at knott.steven@epa.gov.

SAP Minutes and Final Report No. 2017-01

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

EPA's Evaluation of the Carcinogenic Potential of Glyphosate

December 13-16, 2016
FIFRA Scientific Advisory Panel Meeting
Held at the EPA Conference Center
One Potomac Yard
Arlington, Virginia

James McManaman, Ph.D.
FIFRA SAP Chair
FIFRA Scientific Advisory Panel

Date: MAR 16 2017

Steven Knott, M.S.
Designated Federal Official
Office of Science Coordination and Policy, EPA

Steven M. Knott

Date: MAR 16 2017

PANEL ROSTER

FIFRA SAP Chair

James McManaman, PhD
Professor and Chief
Section of Basic Reproductive Sciences
Department of Obstetrics & Gynecology, Physiology & Biophysics
University of Colorado, Denver
Aurora, CO

Designated Federal Official

Steven Knott, MS
US Environmental Protection Agency
Office of Science Coordination & Policy
FIFRA Scientific Advisory Panel
EPA East Building, MC 7201M
1200 Pennsylvania Avenue, NW
Washington, DC 20460
Phone (202) 564-0103
Fax (202) 564-8382
Knott.steven@epa.gov

FIFRA Scientific Advisory Panel Members

Marion F. Ehrich, PhD, DABT, ATS
Professor, Pharmacology and Toxicology
Department of Biomedical Sciences & Pathobiology
Virginia-Maryland College of Veterinary Medicine
Blacksburg, VA

David A. Jett, PhD
Director, National Institutes of Health CounterACT Program
National Institute of Neurological Disorders and Stroke
National Institutes of Health
Bethesda, MD

Joseph Shaw, PhD
Associate Professor
School of Public and Environmental Affairs
Indiana University
Bloomington, IN

Sonya K. Sobrian, PhD
Associate Professor
Department of Pharmacology
Howard University College of Medicine
Washington, DC

FQPA Science Review Board Members

Kenny Crump, PhD
Private Consultant
Ruston, LA

Laura C. Green, PhD, DABT
President and Senior Toxicologist
Green Toxicology LLC
Brookline, MA

Eric S. Johnson, MB; BS (MD), PhD, MPH, DTPH
Professor, Department of Epidemiology
University of Arkansas for Medical Sciences
Little Rock, AR

Barbara L. Parsons, PhD
Research Microbiologist
Division of Genetic and Molecular Toxicology
National Center for Toxicological Research
US Food and Drug Administration
Jefferson, AR

Kenneth Portier, PhD
Vice President
Statistics and Evaluation Center
American Cancer Society
Atlanta, GA

Aramandla Ramesh, PhD
Associate Professor
Department of Biochemistry & Cancer Biology
Meharry Medical College
Nashville, TN

Elizabeth A. (Lianne) Sheppard, PhD
Professor and Assistant Chair of
Environmental and Occupational Health Sciences
University of Washington
Seattle, WA

Emanuela Taioli, MD, PhD
Director, Institute for Translational Epidemiology
Icahn School of Medicine at Mount Sinai
New York, NY

Daniel Zelterman, PhD
Professor of Public Health (Biostatistics)
Department of Biostatistics
Yale School of Medicine
New Haven, CT

Luoping Zhang, PhD
Professor in Toxicology
School of Public Health
University of California, Berkeley
Berkeley, CA

TABLE OF ACRONYMS

ACRONYMS	DESCRIPTION
AAF	2-Acetylaminoflourene
AHS	Agricultural Health Study
AIDS	Acquired Immunodeficiency Syndrome
AOP	Adverse Outcome Pathway
ATS	Academy of Toxicological Sciences
BW	Body Weight
CASAC	Clean Air Science Advisory Committee
CDK	Cyclin-dependent kinase
CI	Confidence Interval
CNV	Gene Copy Number Variation
DABT	Diplomate of the American Board of Toxicology
DNA	Deoxyribonucleic Acid
EFSA	European Food Safety Authority
FACE	Fellow of the American College of Epidemiology
FAO	Food and Agriculture Organization
FDA	Food and Drug Administration
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
FQPA	Food Quality Protection Act of 1996
FRSC	Fellow of the Royal Society of Chemistry
GM	Genetically Modified
HIV	Human Immunodeficiency Virus
HL	Hodgkin's Lymphoma
IARC	International Agency for Research on Cancer
IP	Intraperitoneal
JMPR	Joint FAO/WHO Meeting on Pesticide Residues
MM	Multiple Myeloma
MOA	Mode of Action
MRID	EPA OPP Master Record Identification Number
MTD	Maximum Tolerated Dose
NHL	Non-Hodgkin’s lymphoma
NIOSH	National Institute for Occupational Safety and Health
NRC	National Research Council
NTP	National Toxicology Program
OCSPP	EPA Office of Chemical Safety and Pollution Prevention

ACRONYMS	DESCRIPTION
OECD	Organization for Economic Cooperation and Development
OPP	Office of Pesticide Programs
OR	Odds Ratio
OSHA	Occupational Safety and Health Administration
RR	Relative Risk
SAP	FIFRA Scientific Advisory Panel
SAS	Statistical Analysis System
SCE	Sister Chromatid Exchanges
USDA	United States Department of Agriculture
US EPA or EPA	United States Environmental Protection Agency
WHO	World Health Organization
8-OH-dG	8-hydroxy-2' -deoxyguanosine

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) Scientific Advisory Panel (SAP) has completed the meeting minutes and final report of the SAP meeting regarding scientific issues associated with **EPA's evaluation of the carcinogenic potential of glyphosate.** Advance notice of the SAP meeting was published in the *Federal Register* on July 26, 2016 (81 FR 48794).

Glyphosate is a non-selective, phosphonomethyl amino acid herbicide registered to control weeds in various agricultural and non-agricultural settings. Labeled uses of glyphosate include over 100 terrestrial food crops as well as other non-agricultural sites, such as greenhouses, aquatic areas, and residential areas. Use of glyphosate in the United States and globally has increased over time, particularly with the introduction of glyphosate-resistant crops; however, usage has stabilized in recent years due to the increased number of weed species becoming resistant to glyphosate. Glyphosate is currently undergoing Registration Review, which is a program where all registered pesticides are reviewed at least every 15 years as mandated by the Federal Insecticide, Fungicide, and Rodenticide Act.

Recently, several international agencies have evaluated the carcinogenic potential of glyphosate. In March 2015, the International Agency for Research on Cancer (IARC), a subdivision of the World Health Organization (WHO), concluded that glyphosate was "probably carcinogenic to humans" (Group 2A). Later, in November 2015, the European Food Safety Authority (EFSA) concluded that glyphosate was unlikely to pose a carcinogenic hazard to humans. In May 2016, the Joint Food and Agriculture Organization (FAO) / WHO Meeting on Pesticide Residues (JMPR), another subdivision of the WHO, concluded that glyphosate was unlikely to pose a carcinogenic risk to humans from exposure through the diet.

Recently, EPA collected and analyzed a substantial amount of data informing the carcinogenic potential of glyphosate and utilized its draft ''*Framework for Incorporating Human Epidemiological & Incident Data in Health Risk Assessment*'' (EPA, 2010) to assess its potential carcinogenic hazard. The draft framework provides the foundation for evaluating multiple lines of scientific evidence and includes two key components: (i) Problem formulation and (ii) Use of the mode of action/adverse outcome pathway (MOA/AOP) frameworks. A comprehensive analysis of data on glyphosate from submitted guideline studies and the open literature was performed. This included epidemiological, animal carcinogenicity, genotoxicity, metabolism, and mechanistic studies. Guideline studies were collected for consideration from the toxicological databases for glyphosate and glyphosate salts. A fit-for-purpose systematic review was conducted to obtain relevant and appropriate open literature studies with the potential to inform the human carcinogenic potential of glyphosate. Furthermore, the list of studies obtained from the toxicological databases and systematic review was cross-referenced with recent internal reviews, review articles from the open literature, and international agency evaluations (i.e., IARC, EFSA, and JMPR).

Available data from epidemiological, laboratory animal carcinogenicity, and genotoxicity studies were reviewed and evaluated for study quality and results to inform the human carcinogenic potential of glyphosate. Additionally, as described in the draft ''*Framework for Incorporating Human Epidemiological & Incident Data in Health Risk Assessment,*'' the

multiple lines of evidence were integrated in a weight-of-evidence analysis using the modified Bradford Hill Criteria considering concepts such as strength of association, consistency of observations, dose response, temporal concordance, and biological plausibility.

The focus of this SAP meeting was on soliciting advice from the Panel on the evaluation and interpretation of the available data for each line of evidence and the weight-of-evidence analysis, as well as how the available data inform cancer classification descriptors per the Agency's 2005 *Guidelines for Carcinogen Risk Assessment*. The Agency's evaluation is summarized in an Issue Paper entitled: Glyphosate Issue Paper: Evaluation of Carcinogenic Potential, EPA's Office of Pesticide Programs, September 12, 2016 (EPA, 2016a).

During the FIFRA SAP meeting, US EPA personnel provided the following presentations (listed in order of presentation):

Welcome and Opening Remarks – Jack Housenger, Director, Office of Pesticide Programs

Introduction – Dana Vogel, Director, Health Effects Division, Office of Pesticide Programs

Overview of Glyphosate Registration and Carcinogenic Potential Evaluation – Monique Perron, ScD, Health Effects Division, Office of Pesticide Programs

Systematic Review and Data Collection Methods – Gregory Akerman, PhD, Health Effects Division, Office of Pesticide Programs

Data Evaluation of Epidemiology Studies – Monique Perron, ScD, Health Effects Division, Office of Pesticide Programs

Data Evaluation of Animal Carcinogenicity Studies – Anwar Dunbar, PhD, Health Effects Division, Office of Pesticide Programs

Data Evaluation of Genetic Toxicity – Gregory Akerman, PhD, Health Effects Division, Office of Pesticide Programs

Data Integration and Weight-of-evidence Analysis Across Multiple Lines of Evidence – Monique Perron, ScD, Health Effects Division, Office of Pesticide Programs

PUBLIC COMMENTS

Oral statements:

During the December 13-16, 2016 FIFRA SAP meeting, oral statements were provided by the following individuals and groups.

1) Daniele Court-Marques, MSPS, on behalf of the European Food Safety Authority (EFSA)

2) Lars Niemann, DVM, on behalf of the German Federal Institute for Risk Assessment (BfR)

3) Donna Farmer, PhD, Caroline Harris, PhD, John Acquavella, PhD, James Bus, PhD, Joe Haseman, PhD., David Kirkland, PhD, and Rick Reiss, PhD, on behalf of Monsanto Company

4) James S. Bus PhD, DABT, Fellow ATS, on behalf of Nufarm Americas Inc.

5) Amechi Chukwudebe, PhD, on behalf of BASF Corporation

6) James S. Bus PhD, DABT, Fellow ATS, and Steven Levine, PhD, on behalf of CropLife America

7) Deborah Hommer, on behalf of Virginians for Medical Freedom

8) Scott Slaughter, on behalf of the Center for Regulatory Effectiveness

9) Sabitha Papineni, PhD, on behalf of Dow AgroSciences

10) Jacob Vukich, PhD, on behalf of DuPont Crop Protection

11) Kevin Hoyer, on behalf of the American Soybean Association

12) Andy Hedgecock, on behalf of FMC Corporation

13) Martin Barbre, on behalf of the National Corn Growers Association

14) Amanda Starbuck, on behalf of Food and Water Watch

15) Bill Freese, on behalf of the Center for Food Safety

16) Robert Hamilton, PhD, on behalf of Sumitomo Chemical

17) Montague Dixon, on behalf of Syngenta Crop Protection

18) Michael Hansen, PhD, on behalf of Consumers Union

19) Sheryl H. Kunickis, PhD, on behalf of the US Department of Agriculture

20) Laura E. Mayer, Marghi Barnes, and Kathy Blum, on behalf of Moms Across America

21) Reverend Billy Talen and Ms. Robin Laverne Wilson, on behalf of The Immediate Life Church

22) Nichelle Harriott, PhD, on behalf of Beyond Pesticides

23) Dalia Hashad, PhD, on behalf of Avaaz

24) Peter Infante, DDS, DrPH, FACE, on behalf of himself

25) David Spak, on behalf of Bayer Crop Science

26) Alexis Baden-Mayer, Esq., on behalf of the Organic Consumers Association

27) Luther Markwart, on behalf of the American Sugarbeet Growers Association

28) James Barile, on behalf of the Natural Resources Defense Council

Handouts provided by oral presenters are available in the public docket at https://www.regulations.gov, docket number EPA-HQ-OPP-2016-0385.

Written statements:

Numerous written public comments were submitted to the FIFRA SAP for the December 13-16, 2016 meeting on EPA's evaluation of the carcinogenic potential of glyphosate. These documents are contained in over 350 docket entries and represent the comments of over 260,000 individuals. These comments are available in the public docket at https://www.regulations.gov, docket number EPA-HQ-OPP-2016-0385. Appendix 1 contains a summary list of these docket entries.

EXECUTIVE SUMMARY

US EPA presented a set of charge questions to the FIFRA SAP covering five broad aspects of the Agency's evaluation of the carcinogenic potential of glyphosate. The questions centered on:

1) the completeness, transparency, and appropriateness of the Agency's methods to collect references for the evaluation;

2) the epidemiological studies investigating the potential for associations between glyphosate exposure and cancer outcomes;

3) the laboratory rodent carcinogenicity studies for glyphosate;

4) assays investigating the genotoxic potential of glyphosate; and

5) the completeness, transparency, and scientific quality of the Agency's characterization of the carcinogenic potential of glyphosate for humans.

The completeness, transparency, and appropriateness of the Agency's methods to collect references for the evaluation

The Panel found that EPA's literature review methods were in general transparent and appropriate. However, the Panel provided several recommendations for updated searches that would be more inclusive and capture more recent, relevant publications. In addition, the Panel recommended that the Issue Paper identify and discuss any rodent cancer bioassays of glyphosate-based formulations. Some members of the Panel proposed that searches of "glyphosate and immunotoxicity" and "non-Hodgkin's lymphoma (NHL) and farming" might be informative. Further, some members of the Panel noted that, since most of the glyphosate in commerce in the U.S. is supplied as the isopropylamine salt, it would be of interest to review whether isopropylamine *per se* or the glyphosate isoproylamine combination has been tested for carcinogenicity, mutagenicity, and immunotoxicity.

Given the importance of epidemiologic data generated by the Agricultural Health Study (AHS), the Panel recommended that EPA contact the AHS investigators to determine whether updated data on incidence of non-Hodgkin's lymphoma (NHL) and other cancers are available. As was discussed at length during the Panel's deliberations, the relevant AHS publication (De Roos et al., 2005) has a limited follow-up period, and so is less informative than it might be were additional and more recent data from this important study-cohort available.

One Panel member was concerned with regard to the sensitivity of the review process. The unusually low number of epidemiological studies identified through searches of PubMed.gov, Science Direct®, and Web of Science™ may indicate that EPA needs to utilize more comprehensive and sensitive techniques in conducting searches of the databases than has been employed to date. It is nonetheless likely that the Agency did identify all of the relevant papers by the combined methods of computerized searching and other means (such as from the reference lists of other relevant papers and reviews).

Some Panel members noted that it is important for the study selection process to involve multiple people independently selecting studies, scoring studies, and then to have a process to reach consensus regarding the selected studies. It was noted that this aspect of the process was not clearly described in the Issue Paper.

Several Panel members noted that it would have been helpful if the Issue Paper had been easier to review. For EPA's Clean Air Science Advisory Committee (CASAC), the Agency produces technical documents for review with references linked using HERONET, a database which provides access to full scientific articles. A Panel member suggested that the Agency do the same for FIFRA-related Issue Papers.

The epidemiological studies investigating the potential for an association between glyphosate exposure and cancer outcomes

The Panel concluded that, overall, the Agency's review and evaluation chose relevant epidemiology studies that inform the assessment of the human carcinogenic potential of glyphosate. The Panel noted that EPA's continuing effort to incorporate human data into risk assessment is commendable. The Panel also found that EPA's evaluation of the epidemiologic studies used a sound, appropriate and acceptable approach, although how the individual study rankings were judged and ultimately how the final rankings incorporating subgroup rankings were determined were not always evident to the Panel without the Agency's explanation. In addition, some Panel members were concerned that important issues that affect the quality ranking of the Agricultural Health Study were not considered. The Panel observed that the agency correctly addressed the issue of both case-control and cohort studies having adequate latency periods as a validity criterion, and pointed out the difficulty of addressing this issue in the absence of reliable data on latency periods for the cancers of interest. However, Panelists had different opinions about the importance of considerations of latency in interpreting epidemiology results.

The Panel recommended that the concept of realized study design should be incorporated into the evaluation of study design. In addition, some Panel members suggested that it may be useful to adopt a classification criterion that separates studies by their 1) design, 2) implementation (which includes consideration of issues such as attempts at full enrollment, completeness of questionnaire design, and completeness of collection of other data) and 3) data analyses characteristics.

Panel members agreed that based on the evidence presented in the Issue Paper (EPA, 2016a), Tables 3.3 and 3.4, there is no reliable evidence of an association between glyphosate exposure and any solid tumor, or between glyphosate exposure and leukemia or Hodgkin's lymphoma, even if the possibility that some of the studies reviewed were subject to potential biases is ignored (such as recall or measurement error bias). However, some Panel members also noted that the epidemiologic data are still limited, and that *none* of the studies is of glyphosate manufacturing workers or others who may be relatively highly exposed. This was felt to be a critical data-gap.

The Panel also agreed with EPA that available studies do not link glyphosate exposure to multiple myeloma (MM). However, one Panel member noted that a recently published meta-

analysis (Chang and Delzell, 2016) reported a meta-estimate of the relative risk for the association between MM and glyphosate of 1.4 (with 95% CI of 1.0-1.9). Another panel member, however, noted that to the extent that the primary study results may be biased high, the meta-statistic will be similarly biased high.

Some Panel members supported the Agency conclusion that "the association between glyphosate exposure and risk of NHL cannot be determined based on the available data," although for somewhat different reasons than provided by EPA. These Panelists believe that all the significant findings from three of five case-control studies and three meta-analyses were most likely a result of recall and other potential biases. Furthermore, the only study not subject to recall bias, the prospective cohort study (De Roos et al. 2005), did not show statistical evidence of a positive association.

Some Panel members emphasized that, as EPA itself has estimated, all available epidemiologic studies of glyphosate-users are not really studies of glyphosate over-exposed workers. These Panel members believe this is a crucial point, and one more reason to doubt that the weakly positive NHL case-control study results are indicative of any genuine biological response due to glyphosate -- as opposed to countless other chemical, biological, microbiological, and antigenic factors associated with living or working on a farm. These Panel members noted that many epidemiological studies have reported farmers to be at increased risk of lymphoma (and sometimes leukemia), including decades before glyphosate was used. One Panel member expanded on this noting that while the Agency correctly considered whether studies had adjusted for exposure to other individual pesticides as one of the important criteria for quality assessment, it has not considered the equally important exposure to farm animals (cattle, pigs, sheep, poultry, etc.) that also needs to be adjusted for in determining the quality of epidemiological studies. These animal exposures involve exposure to oncogenic viruses present in the animals, and also to immune system stimulant endotoxins that are particularly of relevance for tumors of the hematopoietic and lymphatic systems, especially as their occurrences predate the introduction of glyphosate and some of the studies reviewed did show them as important risk factors.

Other Panel members disagreed with the Agency's conclusion, emphasizing the value and importance of the findings reported from several dose-response analyses and meta-analyses. These Panelists noted several considerations including that while the majority of the individual studies are not statistically significant, combining the results using meta-analysis shows a scientifically important and statistically significant elevated NHL risk that is relevant for understanding carcinogenic potential. It appeared to some Panel members that the Agency did not fully consider that the data could be suggestive of a lymphomagenic effect of glyphosate. In particular, some Panel members felt that EPA's discussion of the epidemiological evidence appeared to discount statistical findings and overemphasize non-statistical criteria. Thus, some Panel members believed that there is limited but suggestive evidence of a positive association between glyphosate exposure and risk of NHL. These panelists recommended that the Agency revise their conclusion to something along the lines of the following:

"Based on the weight-of-evidence from epidemiological studies and meta-analyses, the Agency cannot exclude the possibility that observed positive associations between glyphosate

exposure and risk of NHL suggest human carcinogenic potential of glyphosate, even though study limitations and concerns about potential biases remain."

Other Panel members, however, strongly disagreed with such a statement; they instead agreed with EPA that the positive associations with glyphosate reported in some retrospective case-control studies of NHL are (i) too weak and (ii) too likely to be confounded by other aspects of living or working on a farm to be properly considered even as suggestive – especially given the null results in the only available prospective cohort study of pesticide applicators. These panelists noted that if the reported odds-ratios and/or relative risks were instead (i) larger and more precise, and (ii) for some solid tumor-type not otherwise known to appear in excess in farmers, then they would be more persuaded that glyphosate possibly posed a cancer-risk. They also noted that if glyphosate, at the very small exposure levels actually received by farmers, were a bona fide human carcinogen, then the toxic potency of glyphosate in humans would have to be on the order of 100,000 times larger than it has proven to be in numerous studies using laboratory rodents. These panelists knew of no precedent for such a discrepancy – especially for a compound, such as glyphosate, that is (i) poorly absorbed, (ii) non-reactive *per se*, and (iii) not converted *in vivo* to reactive metabolites.

Panel members noted that workers in companies that manufacture, formulate, or handle and sell glyphosate on a wholesale basis comprise a promising resource for epidemiologic study that should be investigated. One panel member noted that there are at least 15 companies that have registered glyphosate products with EPA and suggested that it is likely that large numbers of exposed workers (perhaps many more than those directly involved in manufacturing glyphosate) could be identified for cohort studies in companies involved in the formulation or wholesale handling and sale of glyphosate.

The Panel also provided comments and recommendations regarding the specific criteria including study design, study power, statistical analysis, confounding, statistical bias, recall and selection bias. The Panel discussed at length the possibility that recall bias in retrospective case-control studies can result in over-estimation of the risk of NHL associated with pesticide exposure. Some Panel members felt that key studies show evidence of recall bias, exacerbated in some cases by selection bias, and therefore these studies are not reliable for evaluating the carcinogenicity of glyphosate. Other panel members felt that the necessary data to appropriately evaluate whether recall bias is present or not in the reviewed studies are not available and, in any case, the potential for important impacts of recall bias on the findings could not be reliably separated from those of other potential biases. Another Panel member noted, however, that use of proxy respondents (as necessitated in all retrospective case-control studies when cases are deceased) has been shown to bias cancer risk-estimates above the null (sometimes substantially so), both for pesticides in general and for glyphosate in particular.

The laboratory animal carcinogenicity studies for glyphosate

EPA reviewed and analyzed the results of 15 rodent bioassays and concluded that the results as a whole do not indicate carcinogenicity of glyphosate. Some Panel members agreed with this conclusion, noting that the Issue Paper correctly finds the tumor-response data to be too inconsistent to be considered compound-related. Other Panel members interpreted the totality of the tumor data as supporting the hypothesis that glyphosate may cause the promotion or

progression of common spontaneous lesions. These Panel members argued that there is sufficient evidence to conclude that glyphosate is a weak rodent carcinogen and/or tumor promoter. The Panel noted that holistically interpreting results from 15 rodent cancer bioassays posed a unique challenge.

Overall, the Panel was divided with regard to its interpretation of apparently conflicting evidence from the rodent bioassays of glyphosate. Some Panel members pointed out that true carcinogenic responses should be reproducible, and that the estimated positive results in some of the rodent bioassays of glyphosate were likely to be false positives. These Panelists focused on the lack of consistency among the responses across the entire, unusually large glyphosate database, and the fact that the number of significantly positive results in this large database was no greater than would be expected from random assignment of animals to dose groups. These Panelists also noted EPA's weight-of-evidence ignored the serious multiple comparison problem caused by focusing attention on the most extreme tumor responses without also explicitly noting the many negative dose-response relationships and other null results.

Some Panel members felt that the Agency's weight-of-evidence evaluation gave excessive weight to several factors, including lack of monotonic dose response relationships, historical tumor rates, lack of statistical significance in pair-wise comparisons when there is a significant positive trend, and discounting results at exposures greater than the "limit dose" of 1,000 mg/kg/day. Panelists who disagreed with the Agency's conclusions noted there was considerable heterogeneity between studies that needed to be taken into account. They recommended pooled analyses of multiple studies, within endpoint, gender, and species, as a valid approach to distill the evidence from multiple studies. In support of their conclusion they cited an example, provided in the public comments, of pooled analyses of several endpoints for most of the mouse studies.

Some Panel members felt that the Agency's discounting of statistically-significant trends based on the idea that they were not monotonically increasing was flawed. The Panel noted that a monotonic dose response relationship is not a criterion for a positive rodent response in the Agency's 2005 *Guidelines for Carcinogen Risk Assessment.*

Overall, the Panel concluded that the EPA evaluation does not appear to follow the EPA (2005) *Cancer Guidelines* in several ways, notably for use of historical control data and statistical testing requirements. Regarding historical controls, the Panel noted that the default position should be to *not* rely on historical control data except when concurrent controls yield clearly unreliable results. The Panel recommended that EPA articulate why historical control data were incorporated into some of its analyses and not in others. Regarding statistical testing requirements, the Panel noted that requiring a significant pairwise comparison corrected for the number of pair-wise tests in addition to a significant trend is neither consistent with the 2005 *Guidelines for Carcinogen Risk Assessment* nor a conservative approach for public health protection.

In the view of some Panel members, there are sufficient data to conclude glyphosate is a rodent carcinogen using the approaches recommended to interpret the biological significance of tumor responses in EPA's 2005 *Guidelines for Carcinogen Risk Assessment*. However, other Panel members strongly disagreed with this conclusion finding no reliable and consistent

evidence that glyphosate induces or promotes tumors in laboratory rodents. Some Panel members also did not agree that applying a "conservative test" is necessarily an appropriate scientific goal when evaluating the potential carcinogenicity of glyphosate. Instead these Panel members recommended the standard scientific approach be followed whenever feasible (e.g., apply a decision rule that has a false positive rate equal to the standard rate of 5%).

The Panel concluded that the EPA needs to clarify its position on results from exposures that exceed 1,000 mg/kg/day (the limit dose). Panel members differed regarding the relevance and use of results above the "limit dose" for determining the carcinogenic potential of glyphosate for humans. Some Panel members felt that at high doses homeostatic mechanisms could be overwhelmed, so that results might not be relevant for the much lower levels of exposure experienced by people. Other Panelists noted that since glyphosate is so non-toxic, results at dose-rates that are several-fold larger than the limit dose of 1,000 mg/kg/day could indeed be relevant -- since such doses were still smaller than the maximally tolerated dose. Based on EPA (2005) *Cancer Guidelines*, some members of the Panel concluded it is questionable whether results from exposures greater than 1,000 mg/kg/day, but less than doses corresponding to 5% in diet, should be given less weight. Many members of the Panel concluded not considering or discounting tumor responses at doses that exceed 1,000/mg/kg/day is not consistent with either EPA (2005) *Cancer Guidelines* or standard ways in which bioassay results are typically interpreted. They noted that the limit dose is included in the *guidelines* as a design criterion and it is not advisable to exclude observed data *post hoc* from the analysis and interpretation of experimental results.

Some Panel members agreed that it is important to control for multiple comparisons as described in the EPA *Guidelines for Carcinogen Risk Assessment* (a point noted in public comments as well), but felt that the Agency's specific technique for making this adjustment was flawed. These panelists made specific recommendations for improvements in the analysis.

Other Panelists felt that a multiple comparisons adjustment was not appropriate for addressing the question of whether glyphosate has carcinogenic potential, asserting instead that compelling evidence of carcinogenicity for any tumor-type, regardless of replicability, suffices. These panelists felt that the appropriate method for combining evidence from multiple studies is to use pooled analysis or meta-analytical tools.

Some Panel members believed that differences in study designs could explain some of the tumor response discrepancies, and that, overall, the rodent bioassay data were consistent with glyphosate acting as a weak tumor promoter. There has been no direct test of this hypothesis (such as in a standard initiation-promotion bioassay), and therefore other Panel members felt that such a conclusion was speculative and ignored the lack of reproducibility.

Assays investigating the genotoxic potential of glyphosate

Panel members found that the Agency's overall weight-of-evidence and conclusion that there is no convincing evidence that glyphosate induces mutations *in vivo* via the oral route are sound. Areas of remaining uncertainty are related to the potential for glyphosate-induced inflammation and genotoxic effects secondary to toxicity caused by high dose exposures (i.e., glyphosate-induced inflammation, oxidative stress, 8-OH-dG, and sister chromatid exchanges or

SCE) and whether the glyphosate-containing formulations have genotoxic potential. In addition, one Panel member noted that none of the assays employed provides an unbiased (global) measure of small insertions, deletions and rearrangements, which can result in gene copy number variation (CNV) and recommended that this section of the Issue Paper be expanded to address this point.

Panel members agreed that the review and evaluation process of genotoxicity studies is sufficient given the limits of the available assays, which are described in the report (first paragraph of section 5.1) as being sufficient to detect: "1) changes in single base pairs, partial, single or multiple genes, or chromosomes, 2) breaks in chromosomes that result in transmissible deletion, duplication or rearrangement of chromosome segments, and 3) mitotic recombination."

Panel members also agreed that, in the determination of whether glyphosate is likely to be genotoxic in humans, the EPA document focuses appropriately on studies conducted in cultured mammalian cells and laboratory animal models.

One Panel member encouraged the agency to consider two key human biomonitoring studies in their evaluation of genotoxicity, specifically studies by Bolognesi et al. (2009) and Koureas et al. (2014).

A few Panel members commented that if glyphosate causes progression of spontaneously arising lesions (in cells carrying cancer driver mutations or other types of DNA damage), then humans may be at risk of glyphosate-induced carcinogenicity, and the longer human lifespan (as compared to rodents) is expected to contribute to the risk. Other members felt that such concerns were speculative.

The completeness, transparency, and scientific quality of the Agency's characterization of the carcinogenic potential of glyphosate

The Panel was asked to comment on the completeness, transparency, and scientific quality of the Agency's characterization of the carcinogenic potential of glyphosate as presented in the Issue Paper, paying attention to how the Agency uses the modified Bradford Hill criteria of strength of association, consistency, dose response, temporal concordance, and biological plausibility in its assessment.

The Panel noted that the conclusion on glyphosate carcinogenicity offered in the Issue Paper has two parts. The first part is a hazard statement while the second part is a risk characterization statement. Since the Issue Paper is not a full risk assessment of technical glyphosate as outlined in the 2005 *Guidelines for Carcinogen Risk Assessment*, the Issue Paper conclusion was assessed by the Panel as a hazard statement.

Completeness: The Panel concluded that the Issue Paper represents a comprehensive review of the available epidemiologic data, laboratory animal bioassay data, and genotoxicity data, but also noted some limitations.

First, the epidemiologic data reviewed in the Issue Paper are limited to users of glyphosate-based herbicides (such as farmers and other herbicide-applicators), but, as EPA estimates, exposures are fairly low – 0.03-7 mg/kg/day for the most highly exposed workers. Published

studies of potentially more highly exposed workers, such as those who manufacture, formulate or are involved in the wholesale handling or selling of glyphosate, are apparently not available.

Second, because the central epidemiologic question with regard to glyphosate is whether its use is associated with risk of developing non-Hodgkin's lymphoma (NHL), some Panel members felt that the Issue Paper would benefit from a broader review of NHL risk-factors that have long been associated with farming.

Third, the Issue Paper does not present potentially relevant data on isopropylamine, despite the fact that most glyphosate in use is as the isopropylamine salt.

Transparency: The Panel found the Issue Paper to be reasonably transparent, although concern was expressed that some of the documents and data used by EPA in this assessment require special procedures for access and a few studies were not available to the Panel or the public. The Agency explained that FIFRA regulations are responsible for some of these limitations. Regardless, the Panel questioned whether the public could fully review and reproduce the conclusions reached by EPA.

Scientific quality: The Panel felt that the scientific quality of the Issue Paper could be improved. Some Panel members pointed to insufficient study design details, incomplete discussions of data limitations, and use of assessment criteria that do not follow EPA (2005) *Cancer Guidelines*. Panel members noted that the health-effects database on glyphosate (from both toxicological and epidemiological studies) poses a somewhat unique challenge, but that the Agency could nonetheless improve upon the scientific quality of its weight-of-evidence approach. For example, several Panel members, and several public commenters, presented methods for formally and holistically assessing the results from the 15 or so laboratory rodent bioassays of glyphosate acid or glyphosate salts that could improve the Agency's approach.

Dose-response and temporal concordance (Bradford Hill Criteria): A number of Panel members did not agree with how the Issue Paper weighed the epidemiological study findings, particularly for NHL, and were skeptical of the report's arguments leading to its conclusion of "no observed association." Not all Panel members agreed with the Issue Paper's conclusion that findings in rodent bioassays are not treatment-related. There was disagreement among the Panel members regarding which analyses/results constituted a significant finding and which instead were false positives. Some panelists disagreed with EPA's assertion that monotonically increasing dose-response relationships were required in order for responses to be considered to be compound-related, and felt that the Agency could better explain its reliance on tumor responses in historical, as opposed to concurrent, control groups. The Panel's consensus was that the Issue Paper needs to refine and strengthen its arguments regarding the weight assigned to "limit dose" responses in the bioassays. The Panel agreed with the Issue Paper's conclusions regarding the lack of genotoxicity effects of glyphosate.

Strength, consistency, and specificity (Bradford Hill Criteria): With regard to the epidemiologic findings, the Panel concurred with the Issue Paper's conclusions regarding solid tumors, leukemia, multiple myeloma and Hodgkin's lymphoma, but differed in their agreement with the Issue Paper's conclusions of no reliable relationship between glyphosate exposure and NHL. The roles and impacts of recall bias, selection bias, residual confounding by other farm

exposures, and reliability of the meta-analyses were all points of disagreement. Several Panel members noted that the epidemiologic database is *unusually uninformative*, in that (i) glyphosate based herbicide-users are not exposed to doses much larger than those ingested by many consumers via their diets, and (ii) the cancer-type that is weakly associated with glyphosate – NHL – has also been linked with farming for many decades, including before use of this herbicide.

The Panel discussed at length the consistency, or lack thereof, of the laboratory rodent bioassay results. Some Panel members suggest that in evaluating the data from the rodent bioassays, dose-response modeling in a pooled analysis would provide a better basis for assessing the consistency and implications of the bioassay results. The current draft instead focuses on each bioassay individually, which obscures readers' abilities to judge whether results are consistent and likely to be compound-related.

Biological plausibility and coherence (Bradford Hill Criteria): Some Panel members felt that the Issue Paper would benefit from a discussion of the hypothesis that glyphosate may be a weak cancer promoter and to explore the immunotoxicity of glyphosate; though not all Panel members felt that having a biologically plausible MOA is a necessary condition to classifying a substance as a carcinogen, as implied in the Issue Paper. The discussion should consider observations of glyphosate treatment-related increases in frequently occurring spontaneous tumors noted in primary study documents (Knezevich and Hogan 1983, Wood 2009b), observations of treatment-related decreases in pre-neoplastic lesions concurrent with increases in tumor frequency in the same organ (Lankas 1981, Knezevich and Hogan, 1983), and significant increases in malignant tumors of treated male rats relative to controls across tumor sites (Atkinson 1993a), which suggest glyphosate may cause promotion or progression of spontaneous pre-neoplastic lesions (also see response to Charge Question 3d).

Uncertainty (Bradford Hill Criterion): The Panel concluded that uncertainties in epidemiological and animal study evidence are well discussed in appropriate sections of the Issue Paper. Uncertainties identified in earlier sections of the Issue Paper, such as excluding formulations with glyphosate and the limitations regarding available pharmacokinetics data, should be expanded upon. Some Panel members noted that in the discussion of the epidemiology findings, the Issue Paper does not adequately assess the likely impacts of potential biases (such as recall and selection) and residual confounding on the odds ratio estimates or the problems that could bias the estimates obtained from the currently available results of the Agricultural Health Study.

Evaluation and Proposed Conclusion: Using a weight-of-evidence approach, the Issue Paper concludes that glyphosate is "not likely to be carcinogenic to humans," especially at reasonably foreseeable dose-rates. Some Panel members agreed with this characterization, while other Panel members felt that the better descriptor for glyphosate is "suggestive evidence of carcinogenic potential." Many Panelists noted that crucial data were equivocal, and that additional data on cancer morbidity and/or mortality from studies of glyphosate-exposed workers would be desirable.

DETAILED PANEL DELIBERATIONS AND RESPONSE TO CHARGE

To seek advice from the SAP regarding EPA's evaluation of the carcinogenic potential of glyphosate, the Agency presented a set of five charge questions to the Panel focused on the evaluation and interpretation of the available data for each line of evidence, the weight-of-evidence analysis, as well as how the available data inform cancer classification descriptors according to the agency's 2005 *Guidelines for Carcinogen Risk Assessment*. Although there are studies available on glyphosate-based pesticide formulations, the agency solicited advice from the SAP on the evaluation of human carcinogenic potential for the active ingredient glyphosate only at this time.

Charge Question 1

The agency has collected a multitude of studies that may inform the human carcinogenic potential of glyphosate through a systematic review of the open literature and toxicological databases for glyphosate and glyphosate salts as described in Section 2.0. Please comment on the agency's methods to collect references for this evaluation, including the completeness, transparency, and appropriateness of these methods. Please also comment on whether there are additional relevant studies that could inform the human carcinogenic potential of glyphosate that were not included in the current evaluation.

Panel Response

The Panel found that EPA's literature review methods were in general transparent and appropriate. The Panel had the following, additional comments.

First, the Panel questioned whether EPA's search strategy was a bit too narrow. For example, EPA's method of searching excluded papers that contained the term "water" which would have omitted papers reporting on studies of the mutagenic or carcinogenic potential of glyphosate in drinking water. The Panel assumed that this was not EPA's intention, and suggested that updated searches be performed without "water" as an exclusion term.

More generally, the Panel recommended that EPA re-run its literature search to capture recent, potentially relevant papers. Several relevant papers have been published since the Issue Paper was released, and these should be reviewed for the final version of the Issue Paper. These manuscripts include reviews by Acquavella et al., 2016, Williams et al., 2016, and several others that will be readily identified by US EPA when it updates its literature search.

The Panel noted at least one paper that would not be picked up by EPA's current search strategy because glyphosate is not mentioned anywhere in the title and the title does not have any of the other search terms that the Panel believes EPA used. The publication is Zhang et al., 2016, "Health effect[s] of agricultural pesticide use in China: implications for the development of GM crops." Published online in Nature, Scientific Reports (October 10, 2016), the study evaluates what the authors consider to be 35 health indicators in Chinese farmers, and differentiates between users of glyphosate-based formulations and users of non-glyphosate based formulations.

Some Panel members noted that it is important for the study selection process to involve more than one person independently selecting studies for review, more than one person independently scoring the studies, and then to come to some consensus regarding the selected studies. One member thought the Agency said that only one person selected the studies. The selection process is usually done by more than one person and then there is a process to come to some consensus regarding the studies that were selected. Another panelist noted that it's very important that there are at least two people selecting and two people scoring the quality, independently. This member noted that the Issue Paper is not clear regarding this important process.

Given the importance of epidemiologic data generated by the Agricultural Health Study (AHS), the Panel recommended that EPA contact the AHS investigators to determine whether updated data on incidence of NHL and other cancers are available. As was discussed at length during the Panel's deliberations, the relevant AHS publication (De Roos et al., 2005) has a limited follow-up period, and would be more informative if additional and more recent data from this important study-cohort are available.

The EPA's Issue Paper states that it is concerned only with glyphosate *per se*, and not with glyphosate-based formulations. But this is not really the case, since relevant epidemiologic studies would be of people who make or use glyphosate-based formulations.

If there are rodent cancer bioassays of glyphosate-based formulations, the Panel recommended that the Issue Paper identify and discuss these. A Panel member noted that, there is at least one such study, which is the republished bioassay of Roundup® by Seralini's group (Seralini et al., 2014). Some Panelists believed that the EPA should have included and discussed this paper, pointing out its strengths and weaknesses and determining what probative value its reported bioassays provide.

There are other studies that may or may not provide relevant and reliable information, and some members of the Panel wanted to make certain that EPA has evaluated these. Some of these, such as Benedetti and colleagues, 2013, were determined by EPA to be "of low quality ranking and not evaluated in detail," but some members of the Panel suggested that EPA might provide some additional detail as to why they were deemed to be uninformative. Other papers from Seralini's group, such as Mesnage et al (2014) and Cox & Surgan (2006), might also be reviewed to determine if they provide relevant and reliable information. At least one Panel member's review, however, indicated that these papers provide no relevant evidence with regard to the Issue Paper.

Some members of the Panel noted that one of the public comments posted at https://www.regulations.gov (document ID number EPA-HQ-OPP-2016-0385-0235, Attachment 6) makes reference to "Epidemiologic studies from the areas in Latin America where glyphosate is sprayed heavily.…" This comment may refer to work performed by Dr. Fernando Manas at the National University of Rio Cuarto in Argentina who has studied pesticide sprayers in the soy industry. The comment also refers to researchers from the Pontifical Catholic University, Quito, Ecuador who apparently examined individuals exposed when fields were aerially sprayed to eliminate illicit crops. Some members of the Panel suggested that EPA follow up on these potential data sources.

One Panel member noted that there were 18 studies reviewed for which EPA did not have access to the primary reports, and suggested that these be sequestered and considered separately. Another Panel member did not believe that such sequestration would be required.

Several Panel members noted that it would have been helpful if the Issue Paper had been easier to review. For EPA's Clean Air Science Advisory Committee (CASAC), the Agency produces technical documents for review with references linked using HERONET, a database that provides access to full scientific articles. A Panel member suggested that the Agency do the same for FIFRA-related Issue Papers.

Some members of the Panel suggested that additional literature searches might be useful. The central question that arises from the results of epidemiologic studies is whether glyphosate could be a risk factor for the development of NHL. Many types of lymphomas develop at substantially increased rates in patients with compromised immune systems, such as those with HIV-AIDS, and one working hypothesis is that chemicals that are potent immunotoxicants might also predispose to lymphomagenesis. Some members of the Panel proposed that a search of "glyphosate and immunotoxicity" might be informative, as would a separate section in the Issue Paper discussing glyphosate and immunotoxicity test results.

Some members of the Panel also recommended that EPA run a search using the terms "NHL and farming." Studies dating back many decades have often, though not always, reported that farmers develop NHL at excess rates. One working hypothesis for this is that the antigenic stimuli on farms are very different and more diverse than such stimuli in typical non-farm environments. For cancers of lymphocytes – most of which are plasma cell or B-cell neoplasms – immune-system responses are expected to be central to the process of lymphomagenesis.

Finally, some members of the Panel noted that, since most of the glyphosate in commerce in the U.S. is supplied as the isopropylamine salt, it would be of interest to review whether isopropylamine *per se* or glyphosate isopropylamine salt has been tested for carcinogenicity, mutagenicity, and immunotoxicity.

Charge Question 2

As part of its analysis, the agency has considered 58 individual epidemiological studies investigating the potential for an association between glyphosate exposure and numerous cancer outcomes. Detailed study evaluations were performed to determine overall quality rankings for relevant studies. These evaluations took into consideration study characteristics, including study design, exposure assessment, outcome assessment, control for confounders, statistical analyses, and risk of bias. Twenty-three studies were considered informative with regard to the carcinogenic potential of glyphosate.

a. Please comment on the agency's review and evaluation process of relevant epidemiology studies to inform the human carcinogenic potential of glyphosate.

Panel Response

The EPA's Office of Chemical Safety and Pollution Prevention (OCSPP), guided by the NRC recommendation, conducted a systematic review of the epidemiologic data. The Panel found that the review incorporated a transparent and "fit for purpose" approach in identifying high quality studies for selection that was successfully followed in various stages of the review and evaluation process. The Panel noted that EPA's continuing effort to incorporate human data into risk assessment is commendable.

Review Process

EPA initially identified studies for the review from open literature searches of standard databases (PubMed.gov, Science Direct® and Web of Science™). These searches were supplemented with those from other sources such as registrant generated studies submitted to the agency as required under FIFRA, internal reviews and databases, OPP routine evaluations of the epidemiologic literature, evaluations by OPP and other organizations, other governments, and academia. Although on face value it appears a very comprehensive review had been conducted, some members of the Panel found that there is room for concern over the completeness of the review process for the following reasons:

1) The Panel noted that only 9 of the 58 epidemiologic studies selected for review were identified through searches of PubMed.gov, Science Direct® and Web of Science™. Some panel members suggested that this low yield from these sources is quite unusual, and probably indicates a need for the EPA to utilize much more comprehensive, reliable, sensitive, and effective techniques in conducting searches of these databases than has been employed by the agency for this review. One Panel member noted that many of the key papers do not contain glyphosate in their titles or search terms, and so could not have been picked up by computerized searching, hence the need for more innovative methods to capture the relevant studies in these databases. It is possible nonetheless that the Agency did identify all of the relevant papers by the combined methods of computerized searching and other means (such as from the reference lists of other relevant papers).

2) When asked at the meeting, scientists from the Agency revealed that they did not search for nor did they find studies of workers involved in the manufacture of glyphosate for the review. The Panel noted that, historically, for other chemical and physical agents, it has been studies of manufacturing workers that have contributed predominantly in scientific evaluations of the potential carcinogenicity of chemicals and physical agents that pose threats to the general environment and general population (e.g., benzene, aniline dyes, asbestos). Some of the advantages of this group of workers that have been leveraged before in risk assessment include a) they have considerably much higher exposure levels and wider exposure gradients that permit easier detection of effects if any, than users such as applicators and the general population; b) they comprise a well-defined study group that is easily followed up; c) exposures are usually better documented in company/union work histories than in the self-reports associated with population-based or hospital-based case-control studies that are more prone to misclassification of exposure; d) they can be studied in high quality cohort and nested case-control studies that are in principle much better designs than the usual population or hospital-based case-control studies,

especially for issues such as controlling or eliminating selection bias and other confounding factors; and e) occupational exposures may be relatively free of other confounding exposures.

Panel members noted that workers in companies that manufacture, formulate, or handle and sell glyphosate on a wholesale basis comprise a promising resource that should be investigated. One Panel member noted that there are at least 15 such companies that have registered with EPA and suggested that it is very likely that large numbers of exposed workers (perhaps many more than those directly involved in manufacturing glyphosate) could be identified for cohort studies in companies involved in the formulation or wholesale handling and sale of glyphosate. Exposures among such workers would be expected to be much higher than those experienced by applicators. Some Panel members found it is surprising that for this review, the Agency had not requested registrants to provide data on cohort studies they have conducted, as evidence was presented that at least one manufacturing company had conducted such a study. Some Panel members viewed the inclusion of workers involved in manufacturing, formulation or wholesale handling and selling of chemicals such as glyphosate in studies evaluating the carcinogenicity of these chemicals as vital to the review process. This was suggested to be particularly important, since the Agency's entire review process relied on the assumption that applicators have significantly higher exposures than subjects in the general population. Glyphosate exposure of applicators is estimated by EPA to range from 0.02-0.03 mg/kg/day, whereas the EPA's high-end exposure-estimate for children age 1-2 years (assuming that all relevant foods contained glyphosate-residues at their maximum allowable limits) is 0.47 mg/kg/day (i.e., much higher than that for applicators). Even the estimated highest exposures experienced by glyphosate mixers and loaders of 0.03-7 mg/kg/day overlap with those potentially experienced by children. Thus applicators' occupational exposures may not distinguish them from the general population with regard to absorbed doses of glyphosate, and it is not clear then that epidemiologic studies of such users are of much probative value. Members of the Panel urged EPA, OSHA, NIOSH, and industry to collaborate to identify and study workers with distinctly high levels of glyphosate exposure. Because of its importance, the Agency should consider obtaining data on a cohort study of such workers for revision of the Agency's evaluation.

3) Several Panel members noticed that among 58 individual epidemiological studies reviewed, the agency selected a total of 24 human studies to evaluate the human carcinogenic potential of glyphosate (3 rated as high quality and 21 as moderate) not 23, as stated in the Issue Paper.

4) Some Panel members also suggested that the agency should add a cut-off date (e.g. 12-31-2016) to have a newly published and/or accepted paper considered to be included in the EPA 2016 review.

5) For this review, the Scientific Advisory Panel was charged specifically with evaluating the active ingredient, glyphosate acid. However, all the epidemiological studies collected and considered in the review concerned subjects that were exposed to glyphosate formulations, and there are no studies that reflect exposure to glyphosate acid only. This could affect alignment of the epidemiological with the toxicological findings.

Evaluation Process

In the evaluation of epidemiologic studies, EPA tailored study quality considerations specifically to studies investigating the association between glyphosate exposure and cancer endpoints, with primary literature and associated meta-analyses evaluating the association between glyphosate exposure and a cancer endpoint being the focus of the analysis. The EPA judged each study to be of high, moderate, or low quality in each of six domains: study design, exposure assessment, outcome assessment, confounder control, statistical analysis, and susceptibility to bias. The Panel found that this is a sound, appropriate and acceptable approach, although how the individual rankings were judged and ultimately how the final ranking incorporating these subgroup rankings were derived, were not always evident to the Panel without the Agency's explanation. While the classification of studies in the low quality group appeared to be generally appropriate (see the discussion of Cocco et al. in response to charge question 2d), it was not clear how the separation of the three studies in the high quality group differed from others in the moderate group. Several panelists recommended classifying the studies into only two groups because they did not find it clear that studies in the current high quality group could be meaningfully separated from those in the moderate group. These panelists suggested that the "high" and "moderate" quality groups should be combined into a single group, thus reflecting the opinion by some Panel members that the rating should be merely to provide reasonable qualities of the papers included in the evaluation process.

The Panel observed that the agency correctly addressed the issue of both case-control and cohort studies having adequate latency periods as a validity criterion, and pointed out the difficulty of addressing this issue in the absence of reliable data on latency periods for the cancers of interest in the literature. The Panel noted that choice of the "unexposed" group in case-control studies could be a source of differences in findings among them. For example, using non-farmers as the unexposed group could introduce inherent farmer/non-farmer differences unrelated to pesticide exposure that could be confounding, and suggests that the choice of comparison or reference groups may be another quality criterion.

Regarding the Specific Criteria

<u>Study design</u>

The Panel observed that study design is not as clear cut as the document presents. The single cohort of the AHS by De Roos et al. (2005), is given a higher weight than case-control studies, without regard to other extremely relevant aspects of the realized study designs. The Panel recommended that the concept of realized study design should be incorporated into the evaluation of study design. For instance, for multiple reasons, including the young ages of participants, low cancer incidence rate to date, and selection issues, there are important concerns about the AHS, particularly with the published report (De Roos, et al., 2005), that should be taken into account. The usual higher ranking of cohort studies vis-à-vis case-control studies is not applicable in this particular review. Two of the three studies in the high-quality group are from the same AHS cohort, and as mentioned above, this cohort has certain limitations that do not justify its separation into a higher quality ranking over the studies classified as having "moderate quality." Panel members agreed that a follow-up analysis updating results from this cohort could be quite informative.

Some Panel members suggested that it may be useful to adopt a classification criterion that separates studies by their 1) design, 2) implementation (that includes the consideration of issues such as attempts to get full enrollment, completeness of questionnaire design, and completeness of collection of other data), and 3) data analysis characteristics.

Study Power

Members of the Panel observed that study power could have been given too much weight by the Agency. Once a study has been completed, there is no need for further consideration of power. All the evidence is contained in the effect estimate and its confidence interval (CI). The only time Panel members recommend using this criterion is to omit, *a priori*, those studies that have too few cases to estimate adequately the outcome of interest, with the minimum number of cases defined in advance. A member of the Panel suggested that the issue of sample size/power be separated from the statistical analysis assessment in Table 3.2 and moved to study design.

Statistical Analysis

The Panel suggested that the statistical analysis assessment specifically include handling of missing data, adequacy of the analysis models employed, adjustments for confounding, and other characteristics of good, modern data analysis. In addition, the choice of reference group, as justified in the study design and utilized in the data analysis, has important implications for the interpretation of the results, and should be considered in the ranking process.

Confounding

In the report, the Agency stated that the direction of confounding is to inflate any true effect of glyphosate in the absence of statistical adjustment. The Panel noted that this is not always true, and that numerous studies have shown that the effect of confounding can be in either direction (see for example, De Roos et al., 2005; Hoar et al., 1986 and Zahm et al., 1990). The Panel recommended that the discussion not assume the direction of confounding, but consider utilizing bounds on the role of confounders on the effect estimates. This is particularly important for pesticide co-exposures. The Panel recommended that EPA's consideration of the potential carcinogenic effect of other pesticides be addressed in greater detail.

In the report, the Agency correctly notes and uses in its quality assessment whether a study adjusted for exposure to other individual pesticides. It does not consider when assessing the quality of an epidemiological study whether the analysis adjusts for the equally important factor of exposure to farm animals (cattle, pigs, sheep, poultry, etc.). Animal exposures correlate to potential exposures to oncogenic viruses that may be present in the animals, and to immune system stimulant endotoxins that are particularly of relevance for tumors of the hematopoietic and lymphatic systems. The Panel noted that it is well documented that farmers are at increased risk of leukemia and lymphoma, and this risk existed before the introduction of glyphosate. Moreover, some of the studies reviewed by the Agency clearly show statistically significantly elevated risk of NHL in applicators who also reported exposure to certain farm animals. Thus the Panel concluded that exposure to farm animals is equally important as exposures to other pesticides as potential confounders, and that this exposure needs to be accounted for when assessing risk due to glyphosate.

Statistical Bias

Members of the Panel suggested that EPA (2016a) did not consider in its assessment the potential for statistical bias that is likely to occur when fitting models with too many parameters (see a discussion of this in EPA, 2010). As an example, in De Roos et al. (2005), the pesticide-adjusted estimate for the multiple myeloma outcome is based on an analysis model that uses 23 parameters, 15 of these parameters are to account for exposures to other pesticides. This is an excessive number of parameters for a data set with only 32 cases.

Recall and Selection Biases

The Panel discussed these topics at length because some Panel members were concerned that some of the case-control studies may be biased towards showing an effect of exposure to glyphosate due to recall bias and/or selection bias. Selection bias can occur when the controls in a case-control study are not from the same population as the cases. Recall bias can occur if cases tend to consider more carefully than do controls the questions they are asked regarding their exposures or because the cases have been considering what might have caused their cancer (Breslow and Day 1980, Grimes and Schulz 2002). Recall bias is not a problem in cohort studies or in case-control studies nested in cohort studies insofar as these studies ascertain exposure information before cases became diseased (e.g., tumors are diagnosed). However, in all other case-control studies of glyphosate, information on exposure is based on the memories of cases and controls, or their surrogates.

To investigate the potential for recall bias in epidemiological studies of glyphosate, a Panel member constructed a table that tabulates the number of odds ratios (ORs) or relative risks (RRs) greater or less than 1.0 in each of the 18 glyphosate studies EPA considered to be of moderate or high quality. Most of these ORs and RRs are not for glyphosate, as each of these studies evaluated a large number of pesticides. If recall bias were present it would tend to inflate the ORs of all of these pesticides, not just those for glyphosate. This table shows that, overall, there is a large excess of ORs>1 in those 12 studies potentially subject to recall bias. There is a much smaller excess of ORs>1 in the six cohort and nested case-control studies not subject to recall bias, although all of these studies were based on the Agricultural Health Study (AHS) cohort and hence all used the same questionnaire. Nevertheless, this pattern of ORs is exactly what would be expected if recall bias was a significant problem in these studies.

Moreover, this same Panelist noted that the analyses in the case-control studies of NHL by McDuffie et al. (2001), Hardell et al. (2002) and Erickson et al. (2008) eliminated both cases and controls who had been exposed to certain classes of pesticides from their glyphosate-unexposed groups. This analytical choice could cause selection bias, which will tend to exacerbate the effect of any recall bias present. (It would result in unexposed cases being preferentially removed over unexposed controls.). This Panel member conducted a simulation that demonstrated this effect, which also suggested that in certain cases the effect on elevating ORs could be considerable. This convinced some Panel members that the case-control studies of McDuffie et al. (2001), Hardell et al. (2002) and Erickson et al. (2008) likely suffered from selection bias in addition to recall bias, and therefore these studies, in particular, should not be relied upon for evaluating the carcinogenicity of glyphosate.

There are appropriate statistical methods for adjusting for exposure to other pesticides which were used in many of the other case-control studies of glyphosate. These Panel members recommend that before relying on these three studies, EPA attempt to get the data from these studies reanalyzed using an appropriate method.

This analysis does not imply that these or all case-control studies in general suffer from recall bias. In response to the Panelist's analysis described above, other Panelists noted there are many other reasons why this pattern of results might have been found, including an actual effect of pesticides that was not detectable in the AHS due to its design, follow-up, and analysis. (See the following discussion of the AHS for concerns about that study). Some Panelists thought that reliance on memory of remote exposures to pesticides by NHL cases and controls (especially famers) is particularly prone to recall bias while others did not. These other Panelists pointed to the findings discussed in Blair & Zahm (1993) that argue that recall bias by farmers regarding pesticide use could be less likely than for other exposures. One Panelist, in response, presented additional analyses of the data presented by Blair & Zahm showing that the proportion of controls who succeeded in recalling using any pesticides in response to being probed by an interviewer was identical to the proportion of cases with this recall. This identical pattern of recollection suggests no recall bias for pesticide use. Furthermore, in epidemiological studies, recall bias is but one of multiple potential biases that could affect the findings, and the overall impact of these biases on under- or over-estimation of risk (odds ratios) is hard to predict. Some Panel members felt that the necessary data to appropriately evaluate whether recall bias is present or not in reviewed studies is not available, and in any case, the potential for important impacts of recall bias from pesticide exposures on the findings could not be reliably separated from those of other potential biases, especially interviewer bias. Thus the difficulty in adequately quantifying the presence or absence of these biases, their extent, and impact on risk estimates makes it difficult to address adequately their combined impact on study quality.

Another Panel member noted, however, that several of the case-control studies relied on proxy respondents for information regarding use of glyphosate and other pesticides, and that such reliance (although necessary when cases are deceased) has been shown to lead to very substantial overestimation of odds ratios. For example, in a case-control study of brain cancer in Nebraska, Lee et al. (2005) found that the next-of-kin of brain cancer-decedents were much more likely to report that their loved ones had used glyphosate (and other herbicides and pesticides) than did the live cases themselves. In particular, when the analysis was restricted to responses from live cases, the odds ratio for brain cancer and use of any of the class of herbicides that include glyphosate use was 0.7 (95% CI = 0.2-1.8); but when the analysis relied on next-of-kin as proxies, the odds ratio was 3.4 (95% CI = 1.6-7.3). Lee et al. (2005) thus concluded that they had "found significant associations between some specific agricultural pesticide exposures and the risk of glioma among male farmers but not among female farmers in Nebraska; however, most of the positive associations were limited to proxy respondents. These findings warrant further evaluation in prospective cohort studies where issues of recall bias are not a concern."

The Agricultural Health Study (AHS)

The Panel observed that Koutros et al. (2013) is a cohort study and not a case-control study as stated in the Agency's Issue Paper. The AHS design utilizes recruitment of participants from licensed pesticide applicators. Thus the AHS has the advantage of studying a well-defined

population with presumably higher use of pesticides than any other users. However, the Panel was concerned that important issues that impact the quality ranking of this study were not considered.

The AHS utilizes a "prevalent" cohort, in which subjects are not followed from the time of first exposure. Applicators exposed prior to 1993-1997 who did not make it into the study (for various reasons) may have quite different characteristics, including exposure profiles, than those who are included in the prevalent cohort. This is an important point that was not addressed in EPA's evaluation. It has implications for both the exposure-response and the outcome analyses.

The exposure data collected were for the period prior to 1993-1997 when exposures to glyphosate are assumed to have been low. The exposures measured do not adequately capture possibly much higher exposures cohort members likely experienced after the introduction of transgenic crops in 1995. Failure to update exposure data in follow-up implies that the exposure estimates used in the analysis may be, and are most likely significantly underestimated and there would have been misclassification of exposure. Adding to this effect is the fact that the study design precluded the selection of workers with a short latency.

The Panel had other concerns with the AHS. The cohort is relatively young and the follow-up period is brief, both factors limit the time for sufficient events to occur. There is also the issue of statistical bias mentioned above. See additional discussion of the AHS in response to Charge Question 2d.

Summary

Bearing in mind the concerns expressed above, the Panel concludes that, overall, the Agency's review and evaluation chose relevant epidemiology studies that inform the assessment of the human carcinogenic potential of glyphosate.

b. Please comment on the strengths and limitations of the available studies to inform the association between glyphosate and solid tumors, leukemia, and Hodgkin's lymphoma and the agency's conclusion regarding these cancer types described in Section 3.6.

Panel Response

Panel members agreed that based on the evidence presented in the Issue Paper (EPA, 2016a), Tables 3.3 and 3.4, there is essentially no statistical evidence of an association between glyphosate exposure and any solid tumor, or between glyphosate exposure and leukemia or Hodgkin's lymphoma, even if the possibility that some of the studies reviewed were subject to potential biases is ignored (such as recall or measurement error bias). However, some Panel members commented that there were a limited number of available studies to be evaluated in this section and, for some cancer types (e.g., lung, colorectal, breast cancers, etc.), there was only one study available. Therefore, the availability of epidemiologic data is still extremely limited and prevents more in depth discussion of those associations. Additionally, the Panel noted that at the beginning of Section 3.6 (page 63), the Agency mentions one of the 24 epidemiological studies included in the evaluation is uninformative. The Panel requested that the Agency list which study was excluded from the final discussion and conclusion.

Importantly, some Panel members suggested the following points for consideration generally in the reviews of epidemiologic studies: 1) the summaries of all listed relevant studies (Table 3.3) should be expanded to consider topics such as timing of cases and exposure assessment with respect to the registration of glyphosate as well as more details on the exposure assessment; 2) the dose-response summaries should call out comparisons where the referent group was exposed (i.e., the referents were the lowest dose subgroup in the exposed group) or whether there were any exposure lags considered in the analysis; 3) reporting should note the range of risk estimates, as quantified by the range of the CI, and for null effects, including both protective effects and elevated risks, in order to indicate the range of effects consistent with the data; and 4) the discussion should address the conclusions that can be drawn from negative studies. Regarding the last point, Breslow and Day (1987) includes an extensive discussion of the conclusions that can be drawn from negative results in cohort studies. They mention computing confidence limits on estimated excess risk, comparison of the dose levels observed in the study vs. other population exposures, the ability for the risk to have been fully expressed with regard to the time elapsed, the overall risk of the current cohort, and the consistency with other studies.

In summary, the Panel agrees with the Agency's conclusion that there is no evidence of an association between glyphosate exposure and solid tumors, and, there is no evidence of associations between glyphosate exposure and leukemia or Hodgkin's lymphoma. However, the data upon which this evidence is based are very sparse.

c. Please comment on the strengths and limitations of the available studies to inform the association between glyphosate and multiple myeloma. Please comment on the agency's conclusion as described in Section 3.6

Panel Response

The Panel believes there are 4, not 5, studies on multiple myeloma (MM), since Pahwa (2012) and Kachuri (2013) are a re-analysis of the same data set; three case-control and 1 cohort, with a total of 67 exposed cases.

None of the case-control studies report a significant association between MM and glyphosate exposure. None of the case-control studies on MM adjusts for exposure to other pesticides, nor to other aspects of farming that may contribute to risk of developing MM. In addition, Brown et al. (1993) excludes farming cases and controls from the "unexposed" category, which may have introduced selection bias.

De Roos et al. (2005) reports a non-significant suggestion of an exposure-response relationship with regard to glyphosate and MM (P-value = 0.17 for trend with cumulative exposure, based on 19 cases of MM). The Panel agreed with EPA that the imprecision of the risk estimates based on such small numbers of MM cases precludes definitive interpretation. The Panel noted that an updated follow up of this cohort of pesticide-applicators is needed, and is hopefully forthcoming. A Panelist also notes that a reanalysis of the MM data from this study was published subsequent to the EPA document which found that an analysis of the full data set that adjusts for lifestyle factors and exposure to other pesticides produces a reduced OR (1.24, 95% CI: 0.52 to 2.94) over the similar analysis in De Roos et al. based on a reduced data set (OR

= 2.6, 95% CI 0.7 to 9.4), and that removal of adjustment for other pesticides has little effect (Sorahan, 2015).

The Panel agrees with EPA that available studies do not link glyphosate exposure to MM. The Panel notes that a meta-analysis (Chang and Delzell 2016) was published subsequently to the EPA document, and the meta-estimate of the relative risk for the association between MM and glyphosate was 1.4 (with 95% CI of 1.0-1.9).

d. Please comment on the strengths and limitations of the available studies to inform the association between glyphosate and non-Hodgkin's lymphoma (NHL). Please comment on the agency's conclusion as described in Section 3.6.

Panel Response

In the Issue Paper (EPA 2016a; Section 3.6, page 68), EPA states:

"Based on the weight-of-evidence, the agency cannot exclude chance and/or bias as an explanation for observed associations in the database. Due to study limitations and contradictory results across studies of at least equal quality, a conclusion regarding the association between glyphosate exposure and risk of NHL cannot be determined based on the available data. The agency will continue to monitor the literature for studies and any updates to the AHS will be considered when available."

Based on the six primary studies and three meta-analyses reviewed by the Agency (EPA 2016a), Panel members discussed the strengths and limitations of the available epidemiological data. Some panelists noted that the data were limited by the very low exposures received by study subjects.

The Panel could not reach consensus regarding the Agency's conclusion: some of the Panelists agreed with the Agency and others did not. The disagreement among the Panel members is largely due to their different opinions regarding the relative importance they ascribed to potential biases and other challenges with epidemiological studies that could have affected the reported results, as discussed below. Some stressed that NHL in farmers is uniquely difficult to study epidemiologically, given its long-recognized excess apparently due to factors unrelated to glyphosate. This is why the absence of epidemiologic data on glyphosate manufacturers or others who are, (i) not farmers and (ii) likely to be more highly exposed to glyphosate, would be highly desirable, but is apparently absent.

The Panel's detailed response to this question is organized into three sections: Are all original studies selected and rated acceptably? Are the findings from the epidemiological studies described accurately? and Comments on the Agency's Conclusion.

Are all original studies selected and rated acceptably?

The EPA (2016a) identified six epidemiological studies reporting an association between glyphosate exposure and NHL: five retrospective case-control studies and one prospective cohort study. The EPA applied several criteria (Table 3.1) to rate those studies and assigned two of

them (De Roos et al, 2005 and Eriksson et al, 2008) high quality ratings and the remaining four, moderate quality ratings.

Prospective Cohort Study

De Roos et al. 2005 (reporting on results of the Agricultural Health Study, AHS) is the only prospective cohort study, and it received a high rating from the Agency. This study reports on 92 cases of NHL. The relative risk (RR) reported for ever-never exposure adjusted for age, demographic and lifestyle factors, and other pesticides is 1.1 (0.7–1.9). Thus, while there is no evidence of an association in the results from this study, the results are consistent with both a protective effect and an increased risk. There are several challenges with the use of and interpretation of the AHS findings. These include the role of bias, the cohort selection process, the impact of missing data on the results, and the exposure assessment. Even though this cohort was of pesticide applicators, their likely doses of glyphosate were very small, leading to reduced confidence that the results are reflective of glyphosate *per se*, as opposed to myriad other farm-related factors, several of which are known or suspected to be risk factors for NHL.

Impact of Statistical Bias and Missing Data on the Results

DeRoos et al. (2005) used 8 degrees of freedom to adjust for demographic and lifestyle factors for all outcomes. Because the authors did not find evidence of confounding by other pesticides, they did not have an additional 15 degrees of freedom to adjust for other pesticides in their NHL analyses as in their reported MM results. As noted above, there are 92 NHL cases reported in this study prior to exclusions in the adjusted analysis. Thus the risk of statistical bias due to a small number of responses and a large number of parameters in the analysis model is not as strong for NHL as discussed elsewhere for MM. This consideration is based on the *Framework for Incorporating Human Epidemiologic & Incident Data in Health Risk Assessment* (US EPA 2010) that describes statistical bias as going in either direction from models with a large number of parameters and a small sample size (i.e. number of events). Furthermore, it is not known how findings are affected by dropping some cohort members because of missing observations in the adjusted analyses that includes NHL cases in the ever-exposed group. Combining information from Tables 2 and 3 in De Roos, (2005), the Panel noted that it appears 10 exposed cases were dropped from the adjusted analysis (from 71 ever-exposed NHL cases in Table 2 to 61 in Table 3). Some Panel members suggested that a reanalysis be conducted of the full NHL data (i.e. that does not drop cohort members) from De Roos et al. 2005, similar to that conducted by Sorahan 2015 of the MM data from De Roos et al. (2005).

Cohort Selection, Composition, and Follow-Up

The Panel also discussed a concern that the "cross-sectional" enrollment of workers in the AHS could be problematic because it could introduce a survival bias that could bias risk estimates toward the null. This is because pesticide applicators were recruited after their exposure had started and those already diagnosed with cancer were excluded in this prospective study. Another concern is that farmers are known to be at increased baseline risk for NHL. These individuals could be in the reference group in the De Roos et al. (2005) analysis and also exposed to other pesticides, another factor that could increase their baseline risk. The pesticide applicators in the reference group for the glyphosate cohort were most likely exposed to other

pesticides and factors associated with NHL. Thus, even though they addressed adjustment for other pesticides with 15 parameters in their modeling approach, some Panelists thought De Roos et al, (2005) could have underestimated the NHL risk in the ever/never analysis. Some Panel members were concerned that this cohort was followed for an insufficient period of time, and all noted that an updated analysis with additional years of follow up would be informative.

Exposure Assessment

Exposure assessment in the AHS analysis is based on lifetime self-reported pesticide use at baseline, although under- or over-reporting of past exposure is still possible. Using self-administered questionnaires, investigators collected comprehensive use on 22 pesticides and ever/never information on 28 more (De Roos et al., 2005). While the accuracy of self-administered questionnaires relies on individual recall and thorough reporting by respondents, questionnaires are the only feasible tool for exposure assessment in a large cohort such as the AHS. All analyses used one of three exposure metrics: ever/never use, cumulative lifetime use, and intensity-weighted cumulative exposure measured at the time the participant was enrolled in the study. No additional pesticide use is measured after baseline or at any other time during the median of 6.7 years the AHS cohort is followed after enrollment. This poses a challenge to interpretation of the AHS results, if pesticide use varied over time for individual study subjects. The bias introduced by omitting recent exposures is likely to be particularly problematic for the analyses based on cumulative exposure. This omitted exposure bias was only in the direction of under-reporting exposure in the analysis. In other words, many individuals would have been classified in a lower exposure group (including the unexposed group) than was appropriate. These uncertainties can have an important impact on all the dose-response analyses results. Since an exposure lag is typically thought to be appropriate to account for the latency in cancer risk, some Panelists were less concerned about this omission. When exposure lag is taken into account, any exposures within a certain number of years (the lag) before cancer is documented are assumed not to affect cancer risk, where this lag is fixed (and not variable by participant as in the case for the omitted exposure after baseline in the AHS). Although the latency for NHL is unknown, typical exposure lags to account for latency in cancer studies are approximately five years or longer. (See further discussion in the next subsection.)

Furthermore, the Panel noted that the reference group for the dose-response analyses reported in the tables is the lowest tertile of the exposed subgroup, not the more traditional reference being the never-exposed group. De Roos et al (2005) chose to use the lowest exposed tertile in the dose-response analyses in an effort to reduce the potential for residual confounding due to the lack of comparability observed between the never exposed and higher exposed groups. De Roos et al. also conducted analyses using the unexposed group as the reference, and stated that “available data provided evidence of no association between glyphosate exposure and NHL incidence. This conclusion was consistent across analyses using the different exposure metrics and in analyses using either never exposed or low exposed as the referent.”

Latency

In cancer studies, concern about latency (or more accurately, empirical induction as defined by Rothman, 1981) is the time between exposure (or the causal action) and the time of disease diagnosis. Panelists had different opinions about the importance of considerations of latency in

interpreting epidemiology results. For instance, Portier et al. (2016) stated that the follow-up period in De Roos et al. (2005; median = 6.7 years) is not long enough to account for cancer latency and this concern was affirmed by some Panelists who indicated that the median observation period of 6.7 years is too short for a sufficient number of cancer events to develop. (Note that the analysis included only incident cases.) One Panel member pointed out that the exposures in this cohort study reflect the entire exposure of the participants throughout their lives up to the time of their enrollment, and consequently these exposures reflect their total lifetime exposure to glyphosate, except for the additional exposure which occurred between enrollment and the follow-up time used in the analysis. Thus, the time between first exposure and the occurrence of NHL is longer than the follow-up period of 6.7 years for most participants. The Issue Paper (EPA 2016a, Section 3.6, page 67) cited the mean and median exposure durations of 7.5 and 8 years at the time of enrollment, respectively, with a standard deviation of 5.3 years. This inference suggests an exposure duration range of 0-18 years at the time of enrollment, with a higher density of exposures being relatively short (because the mean is smaller than the median). Similarly, counting from the 1974 introduction of glyphosate, the potential latency up to enrollment is 20 years. One Panelist noted that another AHS study (Koutros et al, 2013) indicated that the information on both the amount of exposure per year and the years of exposure was collected by AHS so it would be possible to estimate the time of the initial exposure for all participants. Panelists also noted the evidence presented by Weisenburger (1992) who stated that while median latency for NHL is 5-6 years for high exposures to chemotherapy or radiation, it is expected to be much longer for lower exposures. That paper goes on to state that a median range of 15-20 year latency is plausible for lower chronic exposures. Thus, while some AHS participants would have had exposure durations sufficiently long for a NHL diagnosis to manifest, many were much shorter than the median of 15-20 years.

Additionally, one study by Eriksson et al, 2008 that evaluated the latency effect indicated an increase in NHL risk is related to a longer latency period. This study reports that for latency periods less than 10 years, the NHL odds ratio (OR) = 1.11 (95% CI: 0.24-5.08); and, for latency periods more than 10 years, the OR is increased to 2.26 (95% CI: 1.16-4.40). Hardell et al. (2002) reported an increased risk of NHL for a 10-20 year latency period (called induction period in their paper) of 2.32 (95% CI: 1.04-5.16). Longer induction periods also show evidence of increased risk and are relatively stable over time (OR 1.63, 95% CI: 0.87-2.98 for >20-30 year; 1.70, 95% CI: 1.12-2.58 for >30 year).

Despite the limitations of the De Roos et al. (2005) study discussed by the Panel and agreed upon by most, some Panel members consider this study to be the best of the available epidemiology studies. Panelists who considered the other studies weaker did so based mainly on their opinion that retrospective case-control studies are subject to recall bias. Some Panel members suggested this study should be downgraded from a high to a moderate quality rating or that all the studies should receive the same acceptable quality rating. All Panel members agreed that an update from this cohort study would be most welcomed, now that considerably more person-years have accumulated.

Case-Control Studies

The remaining five studies are all retrospective case-control studies (De Roos et al. 2003, Eriksson et al. 2008, Hardell et al. 2002, McDuffie et al. 2001 and Orsi et al. 2009). Most of the

case-control studies (4 out of 5) are rated moderate quality, with the exception of Eriksson, et al. (2008) that is assigned a high quality rating. Some Panelists argued for a single acceptable rating for all studies. One Panel member questioned whether the Eriksson et al. (2008) study should be given the high rating based on the lack of adjustment for demographic characteristics other than gender and age, and the likelihood of a recall or selection bias. The Panel discussed the Cocco et al. (2013) report on a European multi-center case-control study, to which the EPA assigned a low quality rating and as a result excluded it from consideration in its final evaluation. This study examines associations with the major NHL cellular subtype (B-cell), using confirmation of diagnosis by pathologists. Some Panel members suggested that this study should be included in the NHL evaluation in Table 3.4 (page 61) and its rating be upgraded to moderate (if not high) or "acceptable" quality.

Some Panelists, in view of the evidence cited earlier in response to charge question 2a, considered that each of these case-control studies likely suffered from recall bias, which would tend to bias the ORs in the direction of falsely indicating an effect. This is particularly true of McDuffie et al. 2001, Hardell et al. 2002 and Eriksson et al. 2008, which eliminated subjects exposed to certain classes of pesticides from their unexposed groups, a procedure that will, as explained earlier, exacerbate the effect of any recall bias that may be present.

Based on the same study population as McDuffie et al. (2001), Hohenadel et al. (2011) updated the previous study, corrected four misclassified NHL cases, and reported associations with use of glyphosate with or without malathion. The Panel noted that EPA (2016a) did not include or rate the Hohenadel et al. (2011) study, stating that "This study was not included in the study quality ranking because a more complete analysis was conducted by McDuffie et al. (2001)" in Table 3.2 (page 38). The Panel suggests that EPA should discuss the more complete analysis done in the earlier study and describe in detail how the Agency prioritized the McDuffie et al. (2001) study over the Hohenadel et al. (2011) study. However, a recently conducted meta-analysis by Chang and Delzell (2016) reported that meta-RRs of NHL, regardless of whether they were calculated by including McDuffie et al. (2001) or Hohenadel et al. (2011), were essentially the same with a statistically significantly positive association with glyphosate exposure (meta-RRs=1.3 or 1.4, 95% CI: 1.00-1.6 or 1.00-1.8 from Models 2 or 4).

Additional considerations of case-control studies that apply to this topic, particularly concerns about recall and selection biases, are discussed in response to Charge Question 2a above.

Are the findings from the epidemiological studies described accurately?

EPA (2016a) summarized adjusted effect estimates (RR or OR and 95% CI) obtained from six selected epidemiological studies for NHL in Table 3.4 (pages 61-62). Most of the Panel members agreed that the summary data in this table are informative and were presented reasonably accurately and that the data from the ever/never exposure category were repeated in Figure 3.2 (page 64). However, some important findings are missing in the EPA Report. For example, De Roos et al. (2003) reported an overall effect estimate (OR=2.1, 95% CI=1.1-4.0) for the association between glyphosate exposure and NHL in the standard logistic regression analysis, but the Table 3.4 and Figure 3.2 only showed the results (OR=1.6, 95% CI: 0.90-2.8) from an alternative hierarchical regression analysis. Both these analyses were adjusted for co-

exposure to other pesticides but the hierarchical analysis introduced a prior distribution so the analysis shrunk the estimates towards the overall mean of all other estimates. Several Panel members believed that EPA should not use the hierarchical model estimate in Figure 3.2 (and Table 3.4) for De Roos et al. 2003 but, rather, the standard logistic regression estimate because the standard logistic model estimate is more comparable to the estimates reported by all the other studies.

Some Panelists commented on several incorrect or misleading statements in the discussion of NHL studies in section 3.5.2 (pages 55-58). For example, the EPA states for the De Roos et al. 2005 study (page 56): "study participants provided exposure information prior to enrollment and this information was incorporated into the cumulative lifetime and intensity-weighted cumulative exposure metrics. As a result, the amount of time exposed was longer than just the follow-up time since enrollment." As noted previously, this methodology for reconstructing past exposure is non-optimal for this specific study design, which involved subjects who registered as pesticide users. The procedure of recruiting the cohort from current pesticide applicators and then reconstructing their past exposure introduces "survival bias," since only those who were alive and free of NHL at the time of enrollment had a chance to enter the prospective study. If glyphosate exposure causes NHL, this approach to enrollment would have selectively excluded NHL cases. For Hardell et al. (2002; page 57), EPA states: "The wide range of the CI suggests that the analysis is underpowered." The study is statistically significant, and once a study is complete, all the information about power is contained in the effect estimate confidence interval. Some Panelists suggest EPA instead note that the number of exposed cases was small. For McDuffie et al. 2001, at the end of page 57, EPA states: "It would be expected that effect estimates would attenuate if control for co-exposure to other pesticides had been performed." Several Panelists did not agree with this statement; one cannot say what would happen by adjusting for other pesticides since we do not know if their combined effects are additive, multiplicative, or antagonistic, whether the agent is a promoter, and through what mechanisms these compounds act. Regarding the discussion of Orsi (2009; page 58), EPA concludes: "there is potential for selection bias given the study utilized hospital-based controls". Some Panelists questioned why this indicates selection bias. Panelists also requested that, in general, EPA use the term "adjust" instead of "control."

Several Panelists believed that the discussion about changes in risk over time with increasing use of glyphosate (pages 66-67) is also imprecise and potentially misleading. Similarly, the Issue Paper makes claims about magnitudes of risk estimates for studies in countries with higher use that imply a level of understanding about the information used in the various studies, and the basis for the risk estimates, that is not supported by the information in the document or the original papers. Usage estimates only account for trends in sales, and not trends in the more relevant metric of pounds per worker per year, a quantity that is not known. In terms of incidence, there is only one study from which incidence could be determined, and it is De Roos et al. 2005. Therefore, the comparison of temporal trends in glyphosate usage (sales) with estimated effect estimates is potentially misleading and is not informative for judging the adequacy of the reported NHL results.

Thus, some Panel members recommended: 1) adding the missing positive data in the Table 3.4; 2) correcting the imprecise or misleading statements indicated above; and, 3) providing a

more balanced discussion of the NHL findings in EPA's Section 3.6 (pages 63-69) as discussed below.

Exposure-Response Relationship

The data from the forest plot of effect estimates shown in Figure 3.2 (page 64) indicated that the estimates of NHL risks from ever/never exposure to glyphosate in all epidemiological studies, except Orsi et al. 2009 (OR=1.0, 95% CI=0.5–2.2), were above one (RRs or ORs = 1.10–1.85), though none of the estimates that were included were statistically significant (all 95% CI lower bounds were less than 1.00; see the note above about using the standard logistic regression estimate from De Roos et al. (2003) in place of the hierarchical one included). The Panelists had different perspectives on the interpretation of this observation:

1) Since most of these studies (5 out of 6) were case-control studies, residual bias, including recall bias and other possible biases (such as selection bias, measurement error bias, information bias, and any other uncontrolled/unknown confounders, etc.) could each or in combination be operative in some of them. The EPA (2016a) has detailed the "Risk of Bias" (in 3.2.6, page 29-30) and provided a list of possible biases in the study design summary table (Table 3.2, page 34-43). Based on a weight-of-evidence, the agency concluded that it cannot exclude chance and/or bias as an explanation for observed associations in the database. Some Panel members supported this explanation. Other Panelists believed that in spite of the potential for residual bias in all epidemiological studies and notably in case-control studies, the meta-analysis results (detailed below) are a useful summary of the findings as a whole from the six studies of NHL. Meta-analysis is a common approach of distilling the evidence when only a small number of studies, each with limited statistical power, is available.

2) Some Panel members observed that if the association of glyphosate exposure and NHL did not exist, there should not have been any dose-response relationship detected in any studies by any means. In fact, two studies (Eriksson et al., 2008 and McDuffie et al., 2001) reported an increased risk estimate with increased glyphosate exposure, which is discussed below.

3) In addition, some Panel members observed that, due to a small sample size and/or a limited number of NHL cases identified in each of these studies, individual studies had limited statistical power to detect the association even if it existed. In this case, meta-analysis, combining all of those studies into one analysis to increase statistical power, could be the best method to test this scenario, which is also discussed below.

Potential Biases

As is described at length in the response to charge question 2a, one Panel member detailed the potential risk for recall and other biases and believes that in general, the NHL case-control studies in this review are likely tainted by recall bias, particularly, Eriksson et al. (2008), Hardell et al. (2002) and McDuffie et al. (2001). These three studies are singled out because in their analyses both cases and controls who had been exposed to other pesticides from their glyphosate-

unexposed groups are dropped and this may account for the positive associations seen in these studies. This convinced some Panelists that these case-control studies should not be relied upon for evaluating the carcinogenicity of glyphosate, in spite of them receiving an acceptable quality rating. These Panelists recommended that EPA have the data from these three studies reanalyzed using an appropriate methodology before giving any credence to the results from these studies. Other Panel members were concerned that appropriate data for evaluating whether recall bias was an important factor in any study was not available, and therefore its importance cannot be reliably assessed, and suggested the Panel should not put too much weight on it, and that the Panel should not limit its reanalysis recommendation to a subset of studies. Furthermore, as noted by Panelists and discussed by DeRoos et al. (2005), the results of the study by Blair and Zahm (1993) suggest that there is no evidence of recall bias with respect to pesticide use by farmers. On the contrary, even though no single study could provide sufficient evidence of an absence of recall bias in case-control studies of pesticide exposures, one Panelist presented an analysis of data from Blair and Zahm that indicated statistically significant evidence suggestive of recall bias, in part because the exposure effects of interest were stronger and statistically significant when respondents were probed by interviewers. This claim was contradicted by a follow-up analysis by another Panelist who showed that the proportions of cases and controls who improved their recall was identical. This Panelist also noted that the observed findings would be expected if the extra information attributed to recall bias above could actually be a reduction of measurement error. Further discussion of recall bias is provided in the response to charge question 2a.

Another Panelist noted the challenge of statistical bias is present in many of these studies because there are a large number of parameters used to adjust for potential confounding in many of the studies, thus inflating the uncertainty of the effect estimates of interest and resulting in realized bias of estimates in either direction.

To possibly prevent or limit some of these recall and/or selection biases in case-control studies, another Panel member suggested that if the 'unexposed' category was defined for both cases and controls as not exposed to any pesticide, a true baseline reference could be created for the calculation of the NHL risk with exposure to glyphosate versus non-exposure. Thus, there should be no bias in this approach, since the criterion was applied to both cases and controls. This approach would avoid a higher risk of NHL among the unexposed due to other concurrent exposures that could potentially cause NHL. However, another Panel member pointed out that any recall bias present will be exacerbated by eliminating cases and controls from the unexposed group, because, if recall bias is present, removing cases and controls from the unexposed group will preferentially remove cases over controls. This topic was also noted above and discussed at length in the response to charge question 2a.

Other limitations in the five case-control studies noted by EPA (2016a) include: not adjusting for co-exposure to other pesticides (McDuffie et al. 2001 and Orsi et al. 2009), not adjusting for demographic information (Eriksson et al. 2008, Hardell et al. 2002), the potential for selection bias due to exclusion of observations with missing covariate data (De Roos et al. 2003), and selecting controls from a hospital population (Orsi et al. 2009).

Dose-Response

Among all epidemiological studies, three of six were able to break down exposure into different levels. De Roos et al. (2005) assessed cumulative exposure days and intensity-weighted cumulative exposure days by tertile cut points and this approach did not demonstrate any dose-response relationship between glyphosate exposure and NHL. This study used exposure assessed at enrollment only, and the reference group was the lowest exposed group, as determined from participants' self-reported questionnaire responses at baseline. However, two other studies by Eriksson et al., 2008 and McDuffie et al., 2001 did detect a dose-response relationship. Eriksson et al., 2008 reported that for total glyphosate exposure less than 10 days per year, the NHL OR was 1.69 (95% CI: 0.70-4.07); and, for the total exposure more than 10 days per year, the OR was increased to 2.36 (95% CI: 1.04-5.37), indicating a statistically significant positive association. Similarly, McDuffie et al., 2001 reported the NHL OR was increased from 1 (95% CI: 0.63-1.57) for the use of glyphosate 1 to 2 days per year to 2.12 (95% CI: 1.20-3.73) for the use of glyphosate more than 2 days per year. The same study also found a marginally increased risk of NHL of 1.22 (95% CI: 0.96-1.55) due to higher intensity of any pesticide exposure for more than 10 hours per year versus less.

Despite the evidence of a statistically significant dose-response relationship in both studies (Eriksson et al., 2008 and McDuffie et al., 2001), one Panel member considered that these results should be discounted in view of the potential for bias in these case-control studies and other limitations that were previously discussed. However, other Panelists considered that in case-control studies, a dose-response analysis would be the only way to look at changes (i.e., increase) in individual exposure, even if few human studies have this information. Thus, those Panelists believed that the statistically significant dose-responses were important findings, which not only indicated but also further confirmed the exposure-response relationship and which could not and should not be simply discarded or left undiscussed. Furthermore, some Panelists noted that the exposure quantification and dose-response analyses in De Roos et al. (2005) were biased. In particular, this study systematically undercounted cumulative exposure (because exposure almost certainly continued in this population after baseline, but it was not incorporated into the cumulative exposure metrics) and, unlike the studies to which it was compared, the referent group for this analysis was exposed (at the lowest exposure level; vs. unexposed in the other analyses). Thus, the Panel recommended that EPA (2016a) should at least mention and discuss the importance of dose-response findings in addition to just listing those data in Table 3.4 (pages 61-62).

Finally, as described above, the EPA's claim of an apparent lack of temporal concordance in the risk estimates with glyphosate usage patterns, as discussed on pages 66-67 and again on page 129 of the Issue Paper, could be questioned based on the evidence available. See additional discussion of this topic in the third paragraph of the section "Are the findings from the epidemiological studies described accurately?"

Meta-Analyses

The EPA (2016a) Issue Paper accurately identified three meta-analyses conducted, so far, for the association of glyphosate exposure and NHL, and, all three obtained similar and statistically significant positive results. Schinasi and Leon (2014) reported a meta-effect estimate

of 1.5 (95% CI: 1.1, 2.0) in addition to a positive association between glyphosate exposure and B-cell lymphoma (RR=2.0, 95% CI: 1.1, 3.6), a major subtype (85-90%) of NHL.[1] IARC (2015) modified this analysis by including more fully adjusted effect estimates from Hardell et al. (2002) and Eriksson et al. (2008), and obtained a meta-effect estimate of 1.3 (95% CI: 1.03, 1.65). Using the standard logistic regression results (RR=2.1, 95% CI: 1.1-4.0) from the De Roos et al. (2003) study, Chang and Delzell (2016) obtained a meta-effect estimate of 1.3 (95% CI: 1.0, 1.6) in their model 2 and 1.4 (95% CI: 1.0, 1.8) from their model 4. While the same six epidemiological studies as the EPA (2016a) evaluated were included in the model 2, McDuffie et al. (2001) was replaced by Hohenadel et al. (2011) in model 4 of the meta-analysis. Similarly, McDuffie et al. (2001) in model 1 (meta-RR = 1.3, 95% CI = 1.0-1.6) was also replaced by Hohenadel et al. (2011) in the model 3 (meta-RR = 1.3, 95% CI = 1.0-1.7), but Chang and Delzell (2016) used the alternative hierarchical regression data from De Roos et al. (2003) in their models 1 and 3.

Although the selection of studies included and the criteria for the selection varied across the reported meta-analyses, the data used all originated from the same six studies reviewed by EPA (2016a). To clearly document this point and to precisely display individual studies and specific effect estimates selected in the three meta-analyses, a member of the Panel developed the overview table below (Table 1) to compare the meta-analysis findings with the data included in the EPA 2016a Issue Paper (Figure 3.2, page 64). After taking a close look at Table 1, the Panel realized that the EPA data shown in Figure 3.2 of the Issue Paper (EPA, 2016a) were the same as the data from model 1, shown in Figure 1 of Chang and Delzell (2016) in which the results from the alternative hierarchical regression in the De Roos 2003 study were used. Thus, the meta-effect estimate of 1.27 (95% CI: 1.01, 1.59) shown in Table 1, from the model 1 could be added to Figure 3.2 (page 64), and the Issue Paper (EPA, 2016a) should describe this clearly.

Additionally, some members of the Panel noted that the meta-risk estimates and 95% CI in all four models reported in Chang and Delzell (2016) were reported to one decimal place, particularly all the lower 95% CI equal to 1.0, while some of the original studies used two decimals. Therefore, one Panel member re-analyzed the data from the six original studies and updated those meta-risk estimates and 95% CIs to show two decimal places in Table 1. Evidently, three of the four lower 95% CIs shown are more than 1.0 (Table 1). Thus, these three models from Chang and Delzell (2016), in addition to the two other meta-analyses (Schinasi, 2014 and IARC, 2015), showed a statistically significant positive association of glyphosate exposure and increased NHL risk. However, the EPA Issue Paper (2016a) mistakenly concluded (Page 64, Line 3): "All meta-analysis estimates reported were non-statistically significant except the meta-risk ratio reported by IARC (2015), …." In fact, all three meta-analyses show statistically significant meta-RRs with the lower CI >1.00 (except model 3 in Chang and Delzell, 2016). Some panelists suggested that EPA make this correction and consider adding a new Table similar to the Panel's Table 1 to document meta-risk estimate data from all meta-analyses, including the three published ones and possibly the EPA's own meta-analysis as suggested below. The new table will ideally show two decimals for all CIs and clearly present which

[1] See: https://www.cancer.org/cancer/non-hodgkin-lymphoma/about/types-of-non-hodgkin-lymphoma.html and/or https://www.seattlecca.org/diseases/non-hodgkins-lymphoma/non-hodgkins-lymphoma-facts/types/b-cell-subtypes.

studies were selected and what conditions each of those meta-analyses included (e.g. study selection criteria and assumptions).

The data presented in Figure 3.2, though only based on the ever/never exposure category, showed individual RRs of 1.00–1.85, all with the lower bound of their 95% CIs < 1.00. The Panel observed that in this case, meta-analysis is the best tool to summarize the findings because it, conceptually, uses a statistical approach to combine the results from multiple studies in an effort to increase power (over individual studies), improve estimates of the size of the effect and/or to resolve uncertainty when reports disagree. From multiple sensitivity analyses, all meta-analysis results point to a statistically significant association with the increased risks from 30-50% (meta-RRs=1.3–1.5) for ever exposure to glyphosate. In addition, the meta-analyses consistently show lack of heterogeneity, which speaks for a more robust and credible summary estimate. Some Panelists also suggested that EPA's *post hoc* approach of dividing the studies into the three with higher risks (1.5–1.85) and the three with lower risks (1.00–1.20), in order to argue that the results were contradictory, was not good statistical practice. Meta-analysis is the tool to use to summarize the six study results, particularly in this case where the I^2 statistic indicated low heterogeneity between studies (page 58). The Panel also recommended adding the advantages and reasons of performing meta-analysis in the Issue Paper.

Additionally, some Panel members encouraged EPA to conduct its own meta-analysis and add its meta-RR into the Figure 3.2 forest plot (page 64) for comparison with other published meta-analyses; but other panelists stressed that, since the results from the underlying studies are likely biased high, any meta-statistics would be unreliable. Other Panel members suggested that EPA should conduct a new meta-analysis including all case-control studies (n=5, or n=6 if Cocco, 2013 is included) to prevent heterogeneity of study design and considering the following inclusion criteria: 1) using standard logistic (but not hierarchical) regression; 2) adjusting for other pesticides if available; 3) using the highest exposure dose possible; and 4) using the longest exposure duration possible. The results from this suggested meta-analysis will better address whether or not the exposure to glyphosate in humans, at the highest level and sufficient duration, would increase the NHL risk.

Table 1: Overview of three meta-analyses of glyphosate exposure and NHL, plus individual studies and effect estimates

Studies	Schinasi, 2014	IARC, 2015	Chang & Delzell, 2016[c]				EPA, 2016[d]
			Model 1	Model 2	Model 3	Model 4	
De Roos, 2003	**2.1 (1.1—4.0)[a]**	**2.1 (1.1—4.0)[a]**	1.60 (0.90—2.80)	**2.1 (1.1—4.0)[a]**	1.60 (0.90—2.80)	**2.1 (1.1 — 4.0)[a]**	1.60 (0.90—2.80)
De Roos, 2005	1.1 (0.7—1.9)	1.1 (0.7—1.9)	1.10 (0.70—1.90)	1.10 (0.70—1.90)	1.10 (0.70—1.90)	1.10 (0.70—1.90)	1.10 (0.70—1.90)
Eriksson, 2008	**2.0 (1.1—3.7)[b]**	1.51 (0.77—2.94)	1.51 (0.77—2.94)	1.51 (0.77—2.94)	1.51 (0.77—2.94)	1.51 (0.77—2.94)	1.51 (0.77—2.94)
Hardell, 2002	**3.0 (1.1—8.5)[b]**	1.85 (0.55—6.20)	1.85 (0.55—6.20)	1.85 (0.55—6.20)	1.85 (0.55—6.20)	1.85 (0.55—6.20)	1.85 (0.55—6.20)
Orsi, 2009	1.0 (0.5—2.2)	1.0 (0.5—2.2)	1.00 (0.50—2.20)	1.00 (0.50—2.20)	1.00 (0.50—2.20)	1.00 (0.50—2.20)	1.00 (0.50—2.20)
McDuffie, 2001	1.2 (0.8— 1.7)	1.2 (0.8—1.7)	1.20 (0.83—1.74)	1.20 (0.83—1.74)			1.20 (0.83—1.74)
Hohenadel, 2011					1.40 (0.62—3.15)	1.40 (0.62—3.15)	
Meta-RR (95% CI)	**1.5 (1.1—2.0)**	**1.3 (1.03—1.65)**	**1.27 (1.01—1.59)**	**1.30 (1.03—1.64)**	**1.32 (1.00—1.73)**	**1.37 (1.04—1.82)**	**Not yet estimated**

[a] Data presented as effect-estimate (95% CI), standard logistic regression results reported by De Roos, 2003.
[b] Not adjusted for other pesticides.
[c] Four meta-analyses conducted: Models 1 and 3 used *hierarchical* regression but Models 2 and 4 used standard *logistic* regression results from De Roos, 2003; and Models 1 and 2 included McDuffie 2001 but Models 3 and 4 replaced it with Hohenadel 2011.
[d] Data presented in Figure 3.2 (page 64), used *hierarchical* regression in De Roos, 2003 but not indicating any meta-RR.

Some Panel members noted, however, that reliance on the meta-analyses for NHL should be limited by the likelihood of residual confounding in the original studies. These members noted that NHL in particular has long been noted to be elevated in groups of farmers, including in studies pre-dating use of glyphosate. The unfortunate absence of any epidemiologic results from studies of glyphosate manufacturing workers or others who (i) are not farmers and (ii) were distinctly highly exposed to glyphosate renders any causal interpretation of weakly positive results from the available epidemiologic literature highly problematic. These Panel members also believe that the prospective cohort study of pesticide-applicators, although limited in several important respects, is nonetheless more reliable than the retrospective case-control studies.

One Panel member noted that there is evidence (described in the response to charge question 2a) that all studies entering each meta-analysis in Table 1, except for the cohort study of De Roos et al. 2005, are affected by recall bias, which would tend to cause the ORs from these studies to be biased upward. Moreover, as explained in response to charge question 2a, the nonstandard analyses used in Eriksson et al. 2008, Hardell et al. 2003, and McDuffie et al. 2001 would exacerbate the effect of any recall bias present. Each meta-analysis would be similarly biased, and therefore not reliable for evaluating the carcinogenicity of glyphosate.

Other Panel members believed that since all the studies evaluated for NHL were of acceptable quality and three meta-analyses included by EPA show similar positive meta-RRs with uncertainties suggesting the risk estimates are above 1.0, the evidence from human data is suggestive of the carcinogenic potential of glyphosate. All potential biases described above are plausible, but not sufficient for them to disregard the meta-analyses findings. Thus, those Panel members conclude that there is suggestive evidence of a positive association between NHL and glyphosate exposure, which will be discussed below.

Interpretation and Discussion of Results

Some members of the Panel felt that the discussion of the evidence supporting versus not supporting the NHL findings in EPA (2016a) was highly imbalanced. Despite the fact that evaluation of carcinogenicity in humans always relies on epidemiologic studies, with their own strengths and weaknesses, the EPA's overall discussion appeared to focus on weaknesses and limitations of epidemiology in general as well as in each of the specific studies. It appeared to some Panel members that the Agency did not provide any alternative perspective that the evidence could be suggestive of an underlying effect of glyphosate on NHL. In particular, some Panel members observed that EPA's discussion of the evidence appeared to down-weight statistical findings and up-weight non-statistical criteria. For instance, the discussion of the NHL results uses a non-statistical criterion to classify *post hoc* the effect estimates into two groups based on the size of their point estimate while simultaneously down-weighting the meta-analysis results. Some Panel members felt that meta-analysis is the best tool available to summarize the findings of studies considered to be acceptable by the Agency. Some members of the Panel suggested that it is not good practice to do a *post hoc* division of studies based on effect estimates and use this analysis to discount the evidence. They noted that if the studies are sufficient to be evaluated, then it is most appropriate to use a meta-risk estimate as a summary of the findings. Similarly, given the lack of heterogeneity between studies, and the meaningful overlap in effect estimates across the six NHL studies, it would be inappropriate for EPA to conclude that the studies produced contradictory findings, as was done on page 68.

Some members of the Panel suggested that the Agency also did not appropriately interpret the elevated risks in the context of exposure trends or appropriate understanding of what the effect estimates mean. See the discussion of the Agency's temporal concordance argument above in the third paragraph of the section "*Are the findings from the epidemiological studies described accurately?*"

Comments on the Agency's Conclusion

Regarding the epidemiological studies, the EPA (2006) states (in Section 3.6, page 68): "Based on the weight-of-evidence, the agency cannot exclude chance and/or bias as an explanation for observed associations in the database. Due to study limitations and contradictory results across studies of at least equal quality, a conclusion regarding the association between glyphosate exposure and risk of NHL cannot be determined based on the available data. The agency will continue to monitor the literature for studies and any updates to the AHS will be considered when available."

Supporting the Conclusion

Some Panel members supported the Agency's conclusion above, although for somewhat different reasons than provided by EPA, because they believe that all the significant findings from three of five case-control studies and three meta-analyses were most likely a result of recall bias and other potential biases and confounding. Furthermore, the only study not subject to recall bias, the prospective cohort study (De Roos et al. 2005), did not show statistical evidence of a positive association. Thus, they concurred "the association between glyphosate exposure and risk of NHL cannot be determined based on the available data".

Furthermore, some Panel members put heavy emphasis on the magnitude of the exposure estimates. As these Panelists stressed, and as EPA itself estimated, these studies of glyphosate-users are not really studies of glyphosate over-exposed workers. These panel members believe this is a crucial point, and one more reason to doubt that the weakly positive NHL case-control study results are indicative of any genuine biological response due to glyphosate (as opposed to countless other factors associated with living or working on a farm). These Panelists noted that farmers are not exposed to much more glyphosate than the general population. If the small doses of glyphosate that farmers receive were really carcinogenic, then somehow glyphosate would have to be on the order of 100,000 times more potent in humans than is suggested for mice and rats.

Disagreeing with the Conclusion

Other Panel members disagreed with the Agency's conclusion because they accept the value and importance of the findings reported from multiple dose-response analyses and meta-analyses based on the following points:

1) While the majority of the individual studies are not statistically significant, combining the evidence using meta-analysis shows a scientifically important and statistically significant elevated NHL risk that is relevant for understanding carcinogenic potential. This is based on the lower bound of the meta-risk estimate 95% CIs from all three meta-analyses being consistently greater than or equal to 1.0.

2) Despite the fact that a dose-response of an effect from a specific exposure in human studies can be difficult to detect, two case-control studies reported statistically significant dose-response relationships, which indicated an increased NHL risk estimate with increased exposure. These findings provide further suggestive evidence of the carcinogenic potential of glyphosate.

3) The findings from the collection of NHL studies are not contradictory. In fact, the results are quite consistent and suggestive of a positive carcinogenic potential of glyphosate.

4) Assessing potential bias is a challenge that makes the overall evidence base preliminary; overall the NHL result is suggestive of the carcinogenic potential of glyphosate.

After reviewing all selected studies and the Agency's Evaluation (EPA 2016a), the Panel re-addressed the key question in this evaluation:

Whether or not there is the potential of glyphosate-associated NHL risk in exposed humans?

Overall, some Panel members believed that there is limited but suggestive evidence of a positive association between glyphosate exposure and risk of NHL from epidemiological studies. Therefore, those Panelists concluded and recommended the Agency revise their conclusion to use the following statement:

"Based on the weight-of-evidence from all available data that were abstracted from all qualified human studies, the Agency cannot exclude the possibility of observed positive associations between glyphosate exposure and risk of NHL suggesting human carcinogenic potential of glyphosate even though study limitations and concerns about potential biases remain."

Following the EPA (2005) *Guidelines for Cancer Risk Assessment*, the conclusion of "Suggestive Evidence of Carcinogenic Potential" is considered by some Panel members to be the most appropriate because of the descriptors listed to justify this conclusion. Three of the four guideline examples of when this descriptor is appropriate include:

- If a small, and possibly not statistically significant, increase in tumor incidence observed in a single animal or human study does not reach the weight-of-evidence for the descriptor of "Likely to be Carcinogenic to Humans.

- The study generally would not be contradicted by other studies of equal quality in the same population group or experimental system.

- If there is evidence of a positive response in a study whose power, design, or conduct limits the ability to draw confident conclusions (but does not make the study fatally flawed), but where the carcinogenic potential is strengthened by other lines of evidence.

Some Panel members recommended the following as ways to improve the Issue Paper's discussion of NHL risk:

- Reanalyze the data from all NHL studies using uniform selection and statistical analysis methods.

- Discuss the suggestive association of NHL risk and glyphosate exposure in humans, with the supporting evidence in mouse studies (the positive and monotonic trends of increased lymphomas reported in Wood 2009b, Sugimoto 1997, Knezevich and Hogan 1983 in female mice, as well as Kumar 2001 in male and female mice) and the recent findings on non-genotoxic mechanistic action (e.g. reported from Ford et al., 2017).

Charge Question 3

The agency has followed the 2005 EPA Guidelines for Carcinogen Risk Assessment to evaluate laboratory animal carcinogenicity studies for glyphosate. As described in Sections 4.5 and 4.6, a total of 9 acceptable rat and 6 acceptable mouse carcinogenicity studies were evaluated and considered in the weight-of-evidence analysis. Consistent with the 2005 Guidelines, this analysis took into consideration statistical evidence of a dose-response, the occurrence of corroborating pre-neoplastic lesions or related non-neoplastic lesions to support tumor findings, evidence of progression to malignancy, concurrent and historical control information, and statistical and biological significance of increase tumor incidence, as well as the reproducibility of tumor findings.

a. Please comment on the agency's review and evaluation process of relevant laboratory animal carcinogenicity studies to inform the human carcinogenic potential of glyphosate.

Panel Response

EPA analyzed data from a total of 15 bioassays in mice and rats, noting that many studies showed no compound-related responses, while others showed responses that might or might not reflect compound-related effects. Ultimately, EPA decided that none of the apparently compound-related responses are in fact compound-related, but are instead either false positives or positives only at exceedingly large dose-rates. In discounting these responses, EPA cited some or all of the following factors: 1) lack of monotonically increasing dose-response; 2) absence of pre-neoplastic lesions; 3) incidences in dosed-groups that were within the normal biological variation for a particular tumor type; 4) incidences in the concurrent controls that were not representative of the normal background incidences noted in the historical control animals; and 5) inconsistency in tumor-type-responses in replicate studies.

Panel members disagreed among themselves with EPA's review and evaluation of a total of nine rat and six mice bioassays of glyphosate. In particular, some Panel members focused on individual, statistically significant increases in tumor-responses within individual bioassays, while others focused on the lack of consistency among these responses within the unusually large dataset as a whole. Some Panelists noted that the appropriate approach to evaluate the evidence provided by the collection of findings across multiple studies is to combine endpoint-specific data or results (within gender and species) in pooled analyses, such as meta-analyses. Some panelists suggested that adjustments for different study durations be incorporated in these pooled analyses. It is important that endpoints, species, and genders not be combined in either pooled or meta-analyses because 1) this violates the spirit of the guidelines and 2) the scientific interest is whether there is any carcinogenic potential in any organ that is relevant to humans. All acceptable studies for each outcome analyzed should be reported in the document.

Given the large number of bioassays, the Panel suggested that the EPA Issue Paper might benefit from a holistic presentation and discussion of each tumor-type that appeared to be glyphosate-related in one or more bioassays. This would be consistent with current guidance to use a weight-of-evidence approach to analyze and assimilate all relevant data. This approach mandates use of professional judgment, and may well lead to different conclusions depending on differing but equally well-justified decision rules.

One Panel member considered the Agency's approach for applying the 2005 *Guidelines for Carcinogen Risk Assessment* to the assessment of the glyphosate rodent carcinogenicity data to be flawed, and additional Panelists agreed with this perspective. This Panel member did not find the statistical approaches employed to be consistent with evaluation methods used by other authoritative bodies (e.g., the National Toxicology Program, because the analysis did not include correction for survival) and concluded that at least some of the statistically-significant Cochran-Armitage trend tests and unadjusted pairwise comparisons should be considered to be compound-related (specifically ones that occurred with *P-values* of 0.01 or below).

Many Panel members concluded that the Agency's discounting of statistically-significant trends based on the idea that they were not monotonic was flawed. Regarding statistical evidence of a dose-response, the EPA document discounted four positive tumor responses (tumors with a significant Cochran-Armitage trend test), in part, because the tumor responses were considered non-monotonic. The document discounted three additional positive tumor responses because the dose response was considered shallow. However, the Panel noted that monotonic dose response is not identified as a criterion for a positive rodent response in the EPA's 2005 *Guidelines for Carcinogen Risk Assessment*.

One Panelist noted that it is not good practice to discount the highest dose in these studies, the dose at or above the limit dose, as not relevant to humans. Toxicological studies are designed to detect carcinogenic effects over a range of doses. Sufficiently high doses are needed in order to ensure these studies will be sensitive to the effects of the study compound. The limit dose is used as guidance for study design to determine the highest dose to be studied. This panelist expressed the view that it is not acceptable to selectively discount doses studied in a hazard assessment merely because they are at or above the limit dose. See further discussion of the use of the limit dose in response to Charge Question 3f.

The OECD Test Guidelines 451, 452, and 453 state, "Selection should make provision for survival adjustments, if needed." According to FDA's Guidance for Industry Statistical Aspects of the Design, Analysis, and Interpretation of Chronic Rodent Carcinogenicity Studies of Pharmaceuticals, "the effects of differences in longevity on numbers of tumor-bearing animals can be very substantial, and so, whether or not they (the effects) appear to be, they should routinely be corrected when presenting experimental results." OECD Guidance Document #116, section 4 (page 17) refers to the Cochran-Armitage trend test and states, "Problems arise if there are differences in mortality between the groups. The test is sensitive to increases in treatment related lethality and this leads to an incorrect level of the Type I error (the risk of falsely rejecting the null hypothesis)."

The Glyphosate Issue Paper discusses the lack of reproducibility of significant tumor findings across studies, but did not, in the opinion of many Panel members, provide sufficient discussion of the technical and biological differences that make bioassays performed internationally over a 36-year period unlikely to be true replicates. Also, the Glyphosate Issue Paper does not clearly define the approach or criteria used to identify significant findings across studies. Some Panel members noted a number of places where missing clear definitions/criteria may be responsible for dismissal of significant increases in benign and malignant neoplasms as not being compound-related. Specifically, these panelists noted:

1. The decision to use historical controls did not seem to have been made *a priori* using pre-specified criteria. As a result, the Issue Paper findings appear to have not followed EPA's own guidelines which state, "…statistically significant increases in tumors should not be discounted simply because incidence rates in the treated groups are within the range of historical controls or because incidence rates in the concurrent controls are somewhat lower than average." The Agency used this argument to dismiss increased incidences in the treated groups in several studies. Guidelines also caution against simply relying on a range of the historical responses "… because the range ignores differences in survival of animals among studies and is related to the number of studies in the database." The

suggestion to use historical data gathered within two to three years of the study under review is not followed in evaluating Wood et al. (2009b), since historical controls are from data generated almost a decade earlier from 1987 to 2000. [See EPA (2005) *Guidelines for Carcinogen Risk Assessment* section 2.2.2.1.3., Concurrent and historical controls.]

2. Although the purpose of two or more lower doses is to provide information on the shape of the dose response curve. EPA guidance specifies use of the Cochran-Armitage test to examine whether the results in all dose groups together increase as dose increases. The Cochran-Armitage test is most powerful in identifying significant trend when the underlying trend is linear, but a significant test finding does not imply a linear dose structure, only an increasing response with increasing dose (i.e., an increasing dose-response relationship). In a number of animal studies on glyphosate, the response at doses above 1,000 mg/kg deviated significantly from the near linear trend observed when considering the control and only the remaining lower doses. Decreases in the significance of the tested trend when the highest dose-responses were included in the Cochran-Armitage test could reflect differences in toxicokinetics and/or toxicodynamics at doses above the 1,000 mg/kg dose EPA considered the limit dose for glyphosate.

3. Significant increasing trends as identified by the Cochran-Armitage test, without correspondingly significant pair-wise comparisons test results were also observed with the animal toxicology studies. The Issue Paper discussion often discounts the importance of statistically significant trend test findings when there are no correspondingly significant pair-wise comparisons. This decision rule does not appear to agree with that specified in EPA's guidelines which specify that significance in either of the tests should be considered evidence of significant dose response (i.e., conclude significant response if either Test A OR Test B is significant). While the Sidak adjustment is used to protect against Type I (false positive) indications, this adjustment may also increase the probability of a false negative outcome.

The Agency invoked a four-part "AND" rule: *concluding a significant dose response if one of the pairwise comparisons was significant AND the Cochran-Armitage tests is significant AND the observed trend appears monotonic AND the observed control response rate does not greatly deviate from historical controls levels when available.* Otherwise, one concludes that the dose-response is not significant, as was done in the Issue Paper. These issues lead to the following findings being specified as not treatment related: (77% of the studies cited).

Rats

● Lankas (1981): interstitial cell tumors in testes at 31 mg/kg/day group; pancreatic islet cell adenoma in males; reticulum cell sarcoma in spleen in females.

● Stout & Ruecker (1990): pancreatic islet cell adenomas (males); liver adenomas (males) thyroid cell adenomas (females)

● Brammer (2001): liver adenomas (males)

● Wood et al. (2009a): significant trend for adenocarcinoma and adenoma/adenocarcinoma combined (females).

Mice

● Atkinson et al. (1993b): Hemangiosarcoma in males: Significant trend and significant unadjusted pairwise comparison between the control and high dose groups.

● Wood et al. (2009b): Significant trend for bronchiolar-alveolar adenocarcinoma (males) Significant trend for Malignant Lymphoma (males).

● Sugimoto (1997): Significant trend in hemangioma incidences (females).

Some Panel members noted that the tumor-response data were markedly inconsistent within this unusually large set of bioassay results. For example, at least one Panel member noted and expressed concern the statistically significant report of testicular interstitial tumors (that is, Leydig cell tumors) in one group of Sprague-Dawley rats exposed to glyphosate at a dose-rate of only 31 mg/kg/day was not found important. But another Panel member noted that consideration of the dataset consisting of responses from 8 rat bioassays (Williams et al., 2016) shows there is no relationship between dose and tumor incidence across rat tumor bioassays.

This type of analysis is presented for multiple tumor-types in Williams et al. (2016), and/or could be created by EPA. The Panel recommended that although not a guideline practice, such tabulations, properly analyzed, might form a better basis for EPA's weight-of-evidence evaluation of glyphosate's carcinogenic effects in laboratory rats and mice. Some Panel members while supporting this approach in general expressed doubt that such an analysis would be meaningful, given that the bioassays were performed internationally and over a 36-year time span.

In summary, many Panelists concluded that the Issue Paper's protocol for assessing the significance of laboratory animal carcinogenicity study results does not appear to have followed Agency guidelines. In addition to misinterpreting the rule on assessing significance from combined multiple comparison tests and the Cochran-Armitage trend test, the Issue Paper incorporates into the protocol criteria such as

- exclusion of dose levels considered above the limit dose, without documenting findings that demonstrate that the limit dose was actually exceeded,
- requiring visual confirmation of a monotonic trend in scatter plots of data, which is known to be a poor way of assessing trend, and
- subjectively incorporating information about historical control levels without following other Agency guidance on how and when to incorporate this information.

Many Panelists felt that a more systematic review of the data, organized by endpoint, and that includes in addition to study-specific analyses, a formal statistical analyses of data pooled from all pertinent and quality studies, would result in a stronger assessment.

b. For some of the available animal studies, statistically significant trends in tumor incidence were observed with a lack of statistically significant pairwise comparisons when adjusted for multiple comparisons[2]. Please comment on the agency's methodology and interpretation of statistical analyses to evaluate a linear dose-response (trend test) and increased tumor incidence as compared to controls (pairwise comparisons).

Panel Response

As discussed in Question 3a, many Panelists did not agree with the Issue Paper's methodology and interpretation of statistical analyses to evaluate dose-response and increased tumor incidence as compared to controls. The assessment considered each study separately and required that it show evidence of all of the following (the four-part "AND" rule), *at least one of the Sidak pairwise comparisons was significant AND the Cochran-Armitage test is significant AND the observed trend appears monotonic AND the observed control response rate does not greatly deviate from historical control level when available.*

In this response, the Panel first directly addresses EPA's methodology and interpretation based on the 2005 Guidelines. The Panel suggests alternative approaches to statistical analyses that are not specified in the Guidelines, but that could be considered in future guidelines.

Panel comments directly addressing EPA's use of the 2005 Guidelines for Carcinogen Risk Assessment

According to EPA's 2005 *Guidelines for Carcinogen Risk Assessment*, either a significant trend (Cochran-Armitage) test finding or a statistically significant increase of a treatment group response above its corresponding control group response is sufficient for establishing a finding of a significant treatment-related effect. The document states: "Significance in either kind of test is sufficient to reject the hypothesis that chance accounts for the result" (page 2-19). In requiring that significance be achieved in BOTH tests, the Panel concluded that the Issue Paper's protocol did not follow the 2005 *Guidelines*.

The 2005 *Cancer Guidelines* state that "Considerations of multiple comparisons should also be taken into account. Haseman (1983) analyzed typical animal bioassays that tested both sexes of two species and concluded that, because of multiple comparisons, a single tumor increase for a species-sex-site combination that is statistically significant at the 1% level for common tumors or 5% for rare tumors corresponds to a 7–8% significance level for the study as a whole. **Therefore, animal bioassays presenting only one significant result that falls short of the 1% level for a common tumor should be treated with caution**" (p 2-20).

The remainder of this section addresses three distinct Panel perspectives on the use of the 2005 *Cancer Guidelines* with regard to the large number of studies being reviewed and how multiple comparisons should be handled.

Some on the Panel read the first line of the 2005 Guideline quote above as not mandating the use of multiple comparison tests but leaving its use to the discretion of the analyst. Others on

[2] Individual studies include Stout and Ruecker (1990), Brammer (2001), Wood et al. (2009a), Atkinson (1993b), Wood et al. (2009b), Sugimoto (1997). Results are summarized in EPA's Issue Paper, Table 4.11 and Table 4.18.

the Panel interpreted the line as specifically requiring analysts to account for multiple comparisons (to consider both comparison-wise and experiment-wise errors) in assessing the findings from one or more studies. After extensive discussion, the Panel agreed to recommend that EPA further clarify this statement and make explicit its requirements in this matter.

It was the opinion of some on the Panel that the last line of the 2005 *Guidelines* statement is the driver for a decision rule implemented in the Issue Paper that "down-weights" multiple comparison findings where the associated comparison P-value is greater than 0.01. Logically this rule implies that studies where significant comparisons are observed in multiple tumor types, or where more than one comparison in a single tumor is significant would not have findings "down-weighted" (that is, treated with caution), but would be considered as evidence of a significant dose response. In the Issue Paper, this rule resulted in most study findings being down weighted.

Operationally, by only considering Sidak-adjusted comparisons with a P-value less than 0.01 as significant and worthy of discussion (i.e. as having "weight") a number of findings considered significant by some on the Panel were not discussed. This included:

1) Pancreatic Islet Cell Adenomas and Combined Adenomas and Carcinomas in male Sprague-Dawley rats (Stout & Ruecker: [MRID 41643801])

2) Hemangiosarcomas in Male CD-1 Mice (Atkinson [MRID 49631702])

3) Liver Adenomas in Male Wistar Rats (Brammer [MRID: 49704601]),

4) Mammary gland adenomas/adenocarcinomas combined in female Wistar rats (Wood [MRID: 49957404]),

5) Malignant Lymphomas in Male and Female CD-1 Mice (Wood [MRID: 49957402], and

6) Hemangioma in male and female Specific-Pathogen-Free (SPF) ICR (Crj: CD-1) mice (Sugimoto [MRID: 50017108 and 50017109]).

Thus, in the view of some Panel members, the findings above are sufficient evidence to conclude glyphosate is a rodent carcinogen using the approaches recommended to interpret the biological significance of tumor responses in EPA's 2005 *Guidelines for Carcinogen Risk Assessment*.

Other Panel members, embracing the Issue Paper's use of multiple comparison test controls, concluded that the laboratory animal toxicology studies provide no evidence that glyphosate is a rodent carcinogen. In support of their conclusions, they noted that the animal data on glyphosate encompass 15 bioassays that each employ a control group and three or four treated groups of males and females of rats or mice that combined make up 30 sex-species groups studied. This glyphosate dataset is far more extensive than the NTP bioassay data upon which the Haseman (1983) finding is based. Concern was expressed that the 1% rule may not be adequate for correcting for multiple comparisons in this complex glyphosate data base.

Dr. Haseman, in his presentation to the Panel, showed that the numbers of significant comparison-wise *P-values* at both the 5% and the 1% level in the glyphosate bioassays are no

greater than would be expected purely by chance under the null hypothesis of no treatment effect, when conducting so many statistical tests (i.e., tests from so many different studies by sex by comparisons within tumors).

One Panel member noted that the FDA does not require analysts to use multiple comparison procedures to adjust comparison-wise significance levels but instead sets a small P-value of 0.005 as its cutoff level for establishing significance. There are statistical methods to account for multiple comparisons, and these provide valid statistical tests for use in animal bioassays that properly control the false positive rate (Benjamini and Hochberg 1995, Farrar and Crump 1988, Westfall and Young 1989). Absent results from such tests, the analysis presented to the Panel by Dr. Haseman (discussed above) is another approach for determining whether the number of significant individual comparison tests observed are more or less than would be expected by chance. If there are more than expected, one can conclude that there are statistically significant dose responses in the database, otherwise, one can conclude that there is little evidence.

It was noted that one rat bioassay (Stout and Reucker, 1990) produced three significant tumor observations, which was greater than the number expected by chance. In discussion with the Panel, Dr. Haseman confirmed that the National Toxicology Program has a practice of upgrading (up-weighting) initially non-significant findings (those with a P-value of 0.05 for common tumor types) to "equivocal" when multiple tumors are observed.

The Panel also discussed the importance of consistency of tumor responses across multiple assays as another criterion for assessing the significance of individual outcomes. The Issue Paper handles this issue in an ad hoc fashion, which some Panelists considered appropriate given EPA did not apply formal statistical tests that corrected for multiple comparisons.

To augment EPA's evaluation of consistency among bioassays, one Panel member reported on an analysis they conducted where all of the significant tumor responses in rodent bioassays identified in the issue paper were compared across studies. Responses were compared within the same sex, species and strain of rodent. The only significant tumor response among the ten identified in the Issue Paper that had some support across studies was malignant lymphomas in male CD-1 mice. Wood et al. (2009b) reported a significant dose response trend (P-value=0.0066, high dose ≈ 900 mg/kg/day) as did Sugimoto (1997, P-value=0.016, high dose ≈ 4200 mg/kg/day). However, Wood et al. noted a significant excess response (5/51) at the highest tested dose of 810 mg/kg/day, whereas Sugimoto reported no tumors (0/50) at a comparable dose of 838 mg/kg/day. Knezevich and Hogan (1983, high dose ≈ 5000 mg/kg/day) found no evidence of an effect at any dose tested (P-value = 0.75). Thus the response at the high dose in Wood et al. (2009b) was not reproduced in Sugimoto (1997), and neither the response in Wood nor that in Sugimoto was reproduced in the Knezevich and Hogan study.

After reviewing all the significant tumor responses for consistency, the Panelist who conducted this review concluded that none of the positive findings analyzed in the Issue Paper were reproduced in other studies.

Due to the multiple comparisons evaluations described above, coupled with the lack of consistency among studies, some Panel members concluded that the significant responses that

were observed in the animal studies appeared to best be interpreted as the results of random assignment of animals to dose groups rather than due to any carcinogenic effect of glyphosate. In the view of these Panelists, the interpretation that led to this conclusion is consistent with the 2005 EPA *Cancer Guidelines*. These Panel members also did not agree that applying a "conservative test" is necessarily an appropriate scientific goal when evaluating the potential carcinogenicity of glyphosate. When one uses conservatism of a test as a criterion, there is no clear stopping point, as a more conservative test can always be found. Instead these Panel members recommended the standard scientific approach be followed as nearly as possible (apply a decision rule that has a false positive rate equal to the standard rate of 5%, and otherwise is as powerful as possible).

Representing the third point of view, some Panelists felt that a multiple comparisons adjustment was not appropriate for addressing the scientific question of whether glyphosate has carcinogenic potential from the animal toxicology studies. These Panelists believe that the most important and relevant scientific question is not whether there is compelling scientific evidence of cancer when all cancer endpoints (across all species, and genders) are examined together but **whether there is compelling evidence of carcinogenicity in any one of the endpoints in any single species or gender.** For these Panelist, the analysis should be performed for each cancer endpoint by species by gender combination separately, and data from all relevant studies should be pooled and examined in one omnibus analysis that accounts for study differences. The analysis transmitted as written comments to the Panel by Dr. Christopher Portier, provided a series of examples of such pooled analyses. These are discussed in more detail at the end of the next section.

Additional Panel comments

The Panel also commented more generally on broader statistical issues relevant to the methods and goals described in the 2005 EPA *Guidelines*. It was pointed out that the scientific interest is to assess whether there is a monotonic dose-response trend in the underlying (unobservable) population. Even if the true dose response trend is increasing monotonic, some Panel members suggested that there is no reason to expect a scatter plot of the raw response data to display an increasing monotonic pattern. The Issue Paper confuses the expected underlying rate trend with the observed empirical rate trend. There is a non-negligible probability of observing a non-monotone response when the true underlying response is monotone, primarily due to sampling variability. This was illustrated by one Panel member who presented two realistic situations in which the true dose response was monotone but the observed dose response is non-monotone more often than not. The Panel observed that the non-monotonicity of an observed dose response typically provides very little evidence that the true dose response is non-monotone. Consequently, to some Panel members, the non-monotonicity of the observed dose rate data cannot be used as a valid criterion for down-weighting the results of formal statistical tests that do indicate significant trend.

Unlike pairwise tests, trend tests incorporate all of the dose response data in a single test, and therefore can have greater power for detecting effects than pairwise comparisons. When a trend test and pair-wise comparisons are used together, it could be logically argued the result of the trend test should also be included in the count of multiple comparisons, and the trend test P-value should also be adjusted along with the pair-wise test *P-values* for multiple comparisons.

Thus, conducting both trend and pairwise tests complicates the interpretation of the results. Rather than use multiple tests, some members of the Panel recommend that the Agency consider using a single powerful test for a carcinogenic response, namely a trend test. Although the Cochran-Armitage trend test uses a linear dose-response in its definition, it has power to detect all forms of monotone dose-response. Consequently, when this test is significant, it does not imply that the dose-response is necessarily linear.

A standard Cochran-Armitage trend test could lack power if the observed dose response is non-monotone, as might occur when the highest dose exceeds the maximum tolerated dose (MTD) and causes animals to die prematurely. However, this situation can be allayed by using an age-adjusted trend test, e.g., the poly-3 version of the Cochran-Armitage test.

Finally, logistic, probit, or other linear and non-linear regression methods are more powerful than Cochran-Armitage, but depend on a variety of parametric assumptions. Regression methods offer a simple interpretation as a dose-response. Finally, the Panel encouraged the EPA to explore random effects meta-analysis models and generalized linear models to incorporate multivariate effects of gender, species, and strain to help address multiplicity and increase the power of their findings.

The Panel also had comments on the broad issue of statistical evaluation of pairwise comparisons. Overall, some members of the Panel believe that the EPA over-weighted pairwise comparisons in the Issue Paper. Pairwise comparisons have much lower power than tests for trends. Many published studies in nutrition or epidemiology, for examples, also compare the highest and lowest quartiles, looking for differences in extremes. There are methods to improve on power but they suffer from (1) ignoring much of the data and (2) lack of interpretation.

Finally, the Panel made some comments regarding the issue of multiplicity of statistical tests. The Sidak correction for multiple comparisons is available in SAS routines. It assumes independence. Another widely used correction is due to Bonferroni. For very small *P-values*, the Bonferroni and Sidak provide very similar corrected *P-values*. Sidak is slightly less stringent, i.e. the Bonferroni adjustment will find fewer statistically significant results. The Sidak adjustment is not appropriate when multiple comparisons are not independent, such as when several different groups at different exposures are compared to the same control group. There is also the Dunnett test where the pairwise comparisons of interest are comparisons with a control.

The Benjamini-Hochberg (1995) correction is current state-of-the-art for addressing multiple comparisons and is most generous. This procedure is used in the pharmaceutical industry where the incentive is to identify beneficial properties of new formulations by rejecting the null hypothesis as often as possible and still maintain a constant family-wise error rate.

One Panel member provided a specific numerical example to describe how the statistical analyses are performed. A given set of tables such as the September 9, 2016 memorandum: "Updated Statistics Performed on Animal Carcinogenic Study Data for Glyphosate" gives raw and Sidek-adjusted *P-values* (EPA, 2016b; EPA-HQ-OPP-2016-0385-0095). This is the first pair of tables in the document.

The data taken from Lankas (1981) describes an experiment in which 50 rats each were exposed to one of three different levels of glyphosate added to their diet, including 50 rats given an unexposed diet. All 200 were examined by a pathologist for tumors. Much of the Lankas report was concerned with how much each animal ate as a means of describing the total lifetime exposure. These data and the following statistical analysis were also part of the presentation made by Dr. Dunbar (EPA) on Tuesday, 13 December 2016.

Table 2: Lankas, 1981 (MRID 00093879) - rat Testicular interstitial tumors – males and Corresponding Data Analysis

Exposure	0	3.05	10.3	31.49	Total
No tumor	50	47	49	44	190
Tumor	0 (0%)	3 (6%)	1 (2%)	6 (12%)	10(5%)
Total	50	50	50	50	200

Corresponding data analysis:

Comparison	Test	Raw P-value	Sidak P-value
Four groups	Cochrane-Armitage	0.009	Same
0 to 3.05	Fisher Exact (one tail)	0.121	0.321
0 to 10.3	Fisher Exact (one tail)	0.500	0.875
0 to 31.49	Fisher Exact (one tail)	0.013	0.039

The Cochran-Armitage test compares all four exposure levels and tests for a trend in tumor rates. There is only one Cochrane-Armitage test so there is no adjustment for multiple comparisons. The Fisher exact tests construct three 2x2 tables of frequencies comparing each level of exposure to the unexposed, controls.

In this example, the non-parametric Cochran-Armitage test detects a trend, but only the most extreme of the pair-wise comparisons is statistically significant. The Sidak adjusted P-value corrects for the three pair-wise comparisons.

The three Fisher tests are not independent. Each compares the exposed rats to the same control so the Sidak correction is not valid. The Cochran-Armitage test is also correlated to the three Fisher tests so perhaps there should be an adjustment here for four tests, not three.

The original reference, Lankas (1981), points out the elevated tumor rates given in the table illustrated here and states this several times in their introduction. The pathology report begins on stamped page 2841. There were also female rats in this experiment. All animals were examined for tumors in other body organs under a microscope by a pathologist. Several other indications of tumors were identified but explained. The report lists 32 hematology parameters; 8 organ weights; 38 microscopic examinations for 78 total measurements. Then there were two sexes, 3 doses compared to controls, and an overall trend test for a total of (2 sexes by 3 comparisons by 78 measures + 2 sexes x 78 measures x 1 trend test for a total of) 624 potential tests with associated *P-values*. If we were to restrict our attention to only the 78 trend tests in the Lankas (1981) table, one per measure for one species by sex, and performed each test at the typical type I error rate of 0.05, the chance of making at least one false positive conclusion would be 1-(1-

$0.05)^{77} = 0.98$ – a near certainty. Even if we used instead a much stricter individual comparison type I error significance –value of say 0.009, the chance of making at least one false positive conclusion would be 0.5 (a 1 in 2 chance of saying a comparison is statistically different when in fact they were not).

Not all Panel members agreed that multiple comparisons adjustments should be done to determine the carcinogenic potential of glyphosate or that if done these multiple comparisons should combine all cancer endpoints. Regardless, the number of comparisons is relevant when correcting for multiplicity.

Some Panel members suggested that while not discussed in EPA's (2005) *Cancer Guidelines* as to how it considers the multiple studies for each endpoint, the most appropriate way to address the scientific question at hand -- is there evidence of carcinogenic potential in any endpoint in any species or gender? -- is by conducting a pooled analysis for each species, endpoint, and gender combination. A meta-analysis, such as a random effects meta-analysis, is one possible approach to a pooled analysis. An example for three endpoints in mice, for most of the same studies considered by EPA, was provided in the public comments contributed by Dr. Christopher Portier [EPA-HQ-OPP-2016-0385-0449] and his table is reproduced in this report to demonstrate that the pooled analyses he conducted suggest that there is a carcinogenic potential for some cancer endpoints in at least one animal species. This analysis suggests that EPA's descriptor of "suggestive evidence of carcinogenic potential" is the appropriate descriptor, given that these pooled analyses show compelling statistical evidence of at least one single positive result in at least one species and gender. These Panelists recommend that EPA adopt a pooled analysis approach for combining multiple studies. Adopting a pooled analysis approach should include the development of full guidelines for how to conduct and evaluate these analyses.

Table 3: Meta-analysis as one possible approach to a pooled analysis - Example Provided in public comments contributed by Dr. Christopher Portier [EPA-HQ-OPP-2016-0385-0449]

Study	Tumor	Chi-Squared Test		Exact Test		Historical Control Test		Historical Control
		Original	Poly-3	Original	Poly-3	Original	Poly-3	
Knezevich and Hogan, 1983	Renal Tumors	0.033	0.033	0.063	0.065	0.009	0.009	11/2939
Atkinson, 1993b		0.94	0.937	0.982	0.982	1	1	
Sugimoto, 1997		0.008	0.009	0.061	0.065	0.009	0.002	
Kumar, 2001		0.04	0.044	0.059	0.06	0.011	0.003	
Wood et al., 2009b		0.5	0.5	1	1	0.629	0.797	
All experiments combined		<0.001	0.001	0.003	0.003	0.004	0.007	
All CD-1 Studies Combined		<0.001	0.001	0.005	0.006	0.008	0.008	
All experiments combined, doses<1500		0.212	0.207	0.209	0.206	0.206	0.193	
All CD-1 experiments combined, doses<1000		0.851	0.859	0.856	0.853	0.867	0.906	
Knezevich and Hogan, 1983	Malignant Lymphomas	0.515	0.515	0.736	0.731	0.484	0.478	132/2935
Atkinson, 1993b		0.076	0.076	0.095	0.096	0.087	0.083	
Sugimoto, 1997		0.008	0.012	0.02	0.018	0.013	0.027	
Kumar, 2001		0.053	0.094	0.105	0.105	0.072	0.106	
Wood et al., 2009b		0.004	0.005	0.008	0.008	0.007	0.008	
All experiments combined		0.173	0.193	0.426	0.424	0.172	0.199	
All CD-1 Studies Combined		0.015	0.013	0.084	0.089	0.021	0.023	
All experiments combined, doses<1500		<0.001	<0.001	0.002	0.002	0.001	0.001	
All CD-1 experiments combined, doses<1000		0.031	0.045	0.036	0.036	0.039	0.053	
Knezevich and Hogan, 1983	Hemangiosarcomas	0.628	0.628	0.5	0.504	0.592	0.587	29/2935
Atkinson, 1993b		<0.001	<0.001	0.004	0.004	<0.001	<0.001	
Sugimoto, 1997		0.008	0.009	0.061	0.061	0.021	0.01	
Kumar, 2001		0.724	0.724	0.494	0.5	0.621	0.713	
Wood et al., 2009b		0.5	0.5	1	1	0.49	0.61	
All experiments combined		0.041	0.062	0.056	0.058	0.06	0.08	
All CD-1 Studies Combined		0.024	0.03	0.046	0.047	0.04	0.052	
All experiments combined, doses<1500		0.008	0.006	0.016	0.016	0.015	0.012	
All CD-1 experiments combined, doses<1000		<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	

p<0.01
p<0.05
0.05<p<0.10

c. Unusually low incidences in concurrent controls in comparison with historical controls were noted in Lankas (1981), Stout and Ruecker (1990), and Wood et al. (2009b) and considered as part of the weight-of-evidence for tumor findings. Please comment on the agency's use and interpretation of historical control data as a line of evidence to inform the statistical and biological significance of tumor findings for glyphosate.

Panel Response

Panel comments addressing EPA's use of historical controls vis. a vis. EPA's 2005 Guidelines for Utilizing Historical Control Information

EPA's *Guidelines* for use of historical control data state:

"The standard for determining statistical significance of tumor incidence comes from a comparison of tumors in dosed animals with those in concurrent control animals."

The Panel suggests that this guideline implies that an analysis based on current controls only should normally take precedence over analyses that incorporate historical control data. The *Guidelines* also state:

"Historical control data can add to the analysis, particularly by enabling identification of uncommon tumor types or high spontaneous incidence of a tumor in a given animal strain."

However, the Panel noted that in none of the three studies in which the Issue Paper reports using historical control incidence rates to down-weigh other analysis results were any of these conditions operative. The EPA *Cancer Guidelines* caution against this use of historical control data:

"Generally speaking, statistically significant increases in tumors should not be discounted simply … because incidence rates in the concurrent controls are somewhat lower than average."

The Panel recommends that EPA clearly explain why historical control rates were used in some analyses and not in others. To subjectively choose to use historical control incidence data only in situations where concurrent control incidence levels are low is to potentially introduce biases.

EPA (2016a) considered historical controls in connection with three tumor responses in three studies: testicular interstitial cell tumors in male Sprague-Dawley rats (Lankas, 1981), pancreatic islet cell tumors in male Sprague-Dawley rats (Stout and Ruecker, 1990) and malignant lymphomas in male CD-1 mice (Wood et al., 2009b).

Regarding the induction of pancreatic islet cell adenomas in male Sprague-Dawley rats [Stout and Ruecker, 1990, MRID 41643801] the Issue Paper concludes that the significant tumor response trend was not treatment-related, in part, because the response in concurrent controls was near the lower bound of the historical range. In fact, the control (2.3%) is within the reported historical range (1.8% – 8.3%), while all three treatment groups had a tumor incidence greater than the upper range for the historical control (10% - 18%). Consideration of historical controls

in this case, instead of being used to down-weigh the result, should be used to add to the weight-of-evidence that this is a significant tumor response.

In the response of testicular interstitial cell tumors in male Sprague-Dawley rats [Lankas, 1981, MRID 00093879], the response in concurrent controls was low compared to the range in historical controls (0% versus a historical control range of 3.4% - 6.7%, average 4.5%). The Issue Paper states "Furthermore, the observed incidence of interstitial cell tumors in the glyphosate-treated groups were within the normal biological variation for this tumor type in this strain of rat." The Panel noted that the incidence in the high dose group was 12%, nearly twice the upper bound for historical controls and hence in this case, the historical control data should again add to the weight-of-evidence that this is a significant tumor response.

In the response of malignant lymphomas in male CD-1 mice in Wood (2009b), it is not clear that the difference in the observed concurrent control group incidence (0%) should be considered unusually low, compared to the cited lower bound among historical controls (1.5%).

EPA's (2005) *Cancer Guidelines* mandate careful review of the historical control data to ensure that it is comparable to concurrent data:

"When historical control data are used, the discussion should address several issues that affect comparability of historical and concurrent control data, such as genetic drift in the laboratory strains, differences in pathology examination at different times and in different laboratories (e.g., in criteria for evaluating lesions; variations in the techniques for the preparation or reading of tissue samples among laboratories), and comparability of animals from different suppliers. The most relevant historical data come from the same laboratory and the same supplier and are gathered within two or three years one way or the other of the study under review; other data should be used only with extreme caution."

The importance of using only historical controls from studies conducted closely in time to when the study being evaluated was conducted was highlighted by a discussion one of the Panel members recently had with a breeder with one of the national labs that supplies a large number of rats and mice for these kinds of studies. The breeder pointed out that despite every effort to keep the characteristics of the animals provided as consistent as possible through careful breeding, the genetics of the population drift over time and hence spontaneous cancer rates drift as well. Their conclusion is that the animal breeding populations "drift" much more and faster than human populations do, given the number of generations that can occur over a year of breeding. Therefore, historical control data that is more than three to five years old may not be representative of the animals currently being supplied.

The *Guidelines* also state that:

"Caution should be exercised in simply looking at the ranges of historical responses, because the range ignores differences in survival of animals among studies and is related to the number of studies in the database."

There is no evidence in the Issue Paper that such a careful review was carried out in any of the three studies that utilized historical control information. In the case of Lankas (1981), the Issue Paper reports only the mean and a range for historical control responses that were provided

in the Lankas study report. There is no information on when or where the data/studies were performed from which these historical control values were calculated, hence the relevance of the historical controls is unknown.

In the case of Stout and Ruecker (1990), the Issue Paper provides historical control rates from the same laboratory for seven years (1983 – 1989), apparently from one study for each year. However, there is no indication of the year in which the Stout and Ruecker (1990) study was completed, which raises the possibility that the recommended range of two or three years for historical controls was not met. Also, the historical data reported suggest that control animals that were scheduled for early euthanasia were included, which could have accounted for the lower historical range observed in this study.

In the case of Wood et al. (2009b), the historical control data did not come from the laboratory that performed the study, but from three different laboratories: 59 studies performed during 1987-2000 from two laboratories and 20 studies performed during 1990-2002 from a third laboratory. There is no discussion of whether the procedures used in these laboratories are compatible with those used in Wood et al., which raises the possibility of non-comparability due to different diagnostic criteria, different methods for preparing and reading tissue samples, different times on test, etc. The year of completion of the Wood et al. study is not mentioned, but it appeared to the Panel that the recommendation that only controls from studies completed within the range of two or three years of the completion of Wood et al. could not have been met.

Summary of Evaluation of Agreement of EPA Analysis with EPA Cancer Guidelines

Overall, based on the previous discussion, many Panelists concluded that the use of historical control information in the Issue Paper does not adhere to EPA *Cancer Guidelines*. There is no evidence that the Issue Paper authors performed a careful review of any of the historical control data employed as directed by the EPA *Guidelines* such as discussing the likelihood of genetic drift, differences among animals from different suppliers, differences in laboratory techniques employed in different studies, etc. The timing of studies from which historical control data came is not always clearly stated, although it is clear that the 2 or 3-year limit recommended in the EPA *Guidelines* was not met in some instances. In some cases, the interpretation of results from applying historical controls is questionable. In none of the applications of historical controls was any attempt made to control for survival in the historical controls or even to determine if the survival of the historical controls was compatible with that of the concurrent animals.

Recommendation on EPA Guidelines for use of Historical Controls

The EPA *Guidelines* need to provide more definitive and clearer guidance on when it is appropriate to use historical control information. As currently written the *Guidelines* seem to offer conflicting information, stating “Historical control data can add to the analysis” along with “the standard for determining statistical significance of tumor incidence comes from a comparison of tumors in dosed animals with those in concurrent control animals” and “statistically significant increases in tumors should not be discounted … because incidence rates in the concurrent controls are somewhat lower than average.” A Panel member questioned what the *Guidelines* mean when they recommend using some historical control data only with extreme

caution" or that "caution be exercised in simply looking at ranges…" In addition, it does not quantify what constitutes "unusually low" concurrent control incidence compared to historical control incidence, or whether some statistical test should be used to support this label.

A recent publication discussing the use of historical controls for assessing treatment effects in clinical trials (Viele et al. 2014) offers six methods of incorporating historical control data into testing for treatment effects ranging from ignoring historical data to incorporation of the data using a hierarchical model. Some Panelists recommended that EPA review these and consider adding one or more of these formal methods to its evaluation in the Issue Paper.

Other comments

The EPA document states, drawing from Haseman (1995), that "caution is taken when interpreting results that have marginal statistical significance or in which incidence rates in concurrent controls are unusually low in comparison with historical controls since there may be an artificial inflation of the differences between concurrent controls and treated groups." If this is the case, the significance of comparisons to concurrent controls is still valid. Unusually low control group responses might indicate that this batch of animals are unusually robust, suggesting that they may be unusually resistant to treatment effects. The way the document is written suggests that only the controls are unusually low in comparison to the treatment groups, but if true randomization is used the "unusualness" must apply to the whole batch of animals used.

One Panelist observed that the use of historical controls in the analysis of data can be expressed using Bayesian methods. The Bayesian analysis adds a number of "virtual" animals to the control group and a specified percent of these have tumors. In the analysis of the current data, this is how to express weight to the historical controls.

A member of the Panel also noted that a 2-Acetylaminoflourene (2-AAF) study (Farmer, 1979) used many hundreds of control mice and about 1/3 of 1% of these developed liver cancer. It was suggested that, this information could be used by expressing it as a distribution with corresponding uncertainty or variability in any future study of 2-AAF. Other Panel members felt that too much time has elapsed since this study was conducted for the incidences in controls to be useful in future studies.

Final comment

The Panel agreed with the statement in the EPA *Guidelines* that "The standard for determining statistical significance of tumor incidence comes from a comparison of tumors in dosed animals with those in concurrent control animals." Given this, the Panel believed that the default position should always be to not use historical control information, and, in those cases when it can be used, there should be a clearly articulated reason. When these conditions are satisfied, the requirement in the 2005 *Guidelines* for careful review of the historical data to ensure that it is comparable to concurrent data should be strictly followed.

d. Please comment on the agency's conclusion that there is an absence of corroborating preneoplastic lesions or related non-neoplastic lesions. Please also comment on the agency's conclusion that there is a lack of progression to malignancy to support tumor findings.

Panel Response

The Panel noted that the Issue Paper does not adequately describe its process for identifying and assessing information on pre- and non-neoplastic lesions. From the documentation provided, it was unclear whether a pre-determined process for identifying significant changes in pre- and non-neoplastic findings was employed and, if so, what criteria were used to select data for statistical analyses. For example, it was not clear what kinds of lesions were important for which cancer endpoints in which sexes or species. Also unclear is how the identification of pre-neoplastic lesions supports or is supported by an assumption of a mechanism of tumor induction.

One Panel member noted that in some instances, due diligence in review of the available information was performed and its process included in the Issue Paper. In these cases, further studies were done that included histopathological examination of laboratory animals during and at the end of long-term experiments. Pre-neoplastic lesions should have been noted in tissues collected at midterm euthanasia as well as in tissues collected at the end of the experiment. Observation of pre-neoplastic lesions is expected when some in a group of test animals have neoplastic lesions at the end. The dearth of such pre-neoplastic lesions during and at the end of the experiments presented decreases the likelihood of time-related toxicant-induced effects.

Most Panel members were in agreement that, overall, there is a paucity of evidence for pre-neoplastic findings defined as hyperplasia. When pre-neoplastic lesions were present, in most cases there was no progression to or congruence with adenomas or carcinomas.

Some members of the Panel noted situations where the information reported in the Issue Paper was different than what was reported in the source documents:

1. There are differences in the numbers in tables in the Issue Paper and those in the source documents both with respect to the incidence and number of animals at sacrifice;

2. High dose effects (i.e., >4,000 ppm or 1,000 mg/kg) are discounted despite a lack of pharmacokinetic and/or pharmacodynamic information for doses at this level.

3. There is an increase in the incidence of commonly occurring tumors in aged animals that is not mentioned in the Issue Paper;

4. There are some pre-neoplastic effects attributed to glyphosate treatment in the source documents that are not cited in the EPA Issue Paper;

Panel members found that pre-neoplastic effects attributed to glyphosate treatment in source documents that are not cited in the Issue Paper include the following:

Rat Studies:

● Lankas, 1981: Lymphocytic hyperplasia increased in the thymus of males (M) and females (F), in the mediastinal lymph node (M), the mesenteric lymph node (F), and the spleen (M), in addition C-cell hyperplasia (F) was observed in the thyroid. Spleen reticulum cell sarcoma was reported in females.

● Atkinson et al. 1993a: The most notable histological finding was seen in the salivary glands, where cellular alteration was seen in the submaxillary and/or parotid of males and females.

● Brammer, 2001: In kidney, slight increased incidence of papillary necrosis was observed in M>F, with varying degrees of mineralization of the papilla and/or transitional cell hyperplasia. In pancreas, exocrine hyperplasia increased in high-dose males (within historical control values).

● Suresh, 1996: In spleen, lymphoid hyperplasia increased in the low dose group. Specifically, lymphoid hyperplasia of the mediastinal lymph nodes increased in the low dose group and increased in the mandibular lymph nodes of low and high dose groups. In liver, clear cell foci/areas increased in low and mid dose groups and were described as "Maybe pre-neoplastic changes as evidenced by the high incidence of hepatic tumors in these animals," in the primary study document.

● Wood et al, 2009a: In kidney, hyperplasia of the pelvic/papillary epithelium was observed in the high dose group. In males, pancreas islet cell hyperplasia (minimal) increased in the mid dose group [2/12, 4/13, 7/13, 0/6] and thyroid C-cell hyperplasia increased in the high dose group.

Mouse Studies:

● Knezevich and Hogan, 1983: In liver of males, focal hyperplasia and testes interstitial cell hyperplasia were observed in the low dose group. In uterus, cystic endometrial hyperplasia was observed.

● Atkinson et al., 1993b: The incidence of lung masses was slightly higher in the high dose group (18/50) compared to control (10/50).

Most Panel members were in agreement that, overall, there is a paucity of evidence for dose-related increases in pre-neoplastic findings co-occurring with increases in adenomas or carcinomas in the same tissue. However, significant lymphoid hyperplasia was observed at low and mid doses in males (71.4 and 234.2 mg/kg/day) in a study where malignant lymphomas were significantly induced at 810 mg/kg/day (Wood, 2009a).

It was noted that two studies identified an inverse relationship between dose and the incidence of pre-neoplastic lesions. In the Atkinson et al. (1993a) study of Sprague-Dawley rats, a significant decrease in kidney hyperplasia was observed in female rats (total incidence score: 18/50, 17/49, 13/50, 9/49, and 1/50 for doses of 0, 10, 100, 300, and 1000 mg/kg-bw/day). In the Knezevich and Hogan (1983) study of CD-1 mice, "There was actually a decrease in renal tubule epithelial changes (basophilia and hyperplasia in males, and although there was a dose-related increase in these changes in females, no tubular neoplasms were observed in females)." Some Panel members concluded that these observations are consistent with the interpretation that glyphosate is non-genotoxic and does not cause *de novo* pre-neoplastic lesions, but rather glyphosate is a weak, non-genotoxic carcinogen that causes outgrowth of pre-existing spontaneous lesions. This interpretation is supported by multiple positive tumor bioassays [significant in individual or combined analyses (see Charge Question 3b), as well as three positive tumor responses observed in a single bioassay], the lack of evidence of genotoxicity,

treatment-related increases in frequently occurring spontaneous tumors (Knezevich and Hogan 1983, Wood 2009b), treatment-related decreases in pre-neoplastic lesions concurrent with increases in tumor frequency within the same organ (Lankas 1981, Knezevich and Hogan, 1983), and significant increases in malignant tumors of treated male rats relative to controls across tumor sites (Atkinson 1993a).

Regarding the Issue Paper's conclusion that there is a lack of progression to malignancy to support tumor findings, one Panel member noted that three different malignancies were significantly induced by glyphosate ($P < 0.01$ for malignant hemangiosarcoma, lung adenocarcinoma, and malignant lymphoma). In addition, the study of glyphosate-treated Sprague-Dawley rats by Atkinson et al. (1993a) stated "the overall number of animals with tumors was similar between groups (44/50 in control vs 41/50 in high dose males: 49/50 in control and high dose females), but the number of males in the high dose group with malignant tumors (15/50 or 30%) was almost double that observed in controls (8/50 or 16%)." The Wood study (2009b) of CD-1 mice reported an overall increase in multiple malignant tumors in treated males relative to controls. For this Panel member, these observations regarding malignancy add to the weight-of-evidence that glyphosate is a rodent carcinogen.

The Panel recommends that the Issue Paper revise and strengthen its discussion on the process used for identifying and assessing information regarding pre- and non-neoplastic lesions. In the process, the Issue Paper should address the exclusions discussed above.

e. In the case of glyphosate, there are multiple carcinogenicity studies available for the evaluation of carcinogenic potential. The agency looked across all of the studies and found that tumor findings were not consistent or reproduced in other studies conducted in the same species and strain at similar or higher doses. Please comment on the interpretation of conflicting evidence and reproducibility for these studies.

Panel Response

The Panel was divided with regard to the interpretation of apparently inconsistent evidence from the rodent bioassays of glyphosate. Some Panel members pointed out that true carcinogenic responses should be reproducible, and that the apparently positive results in some of the rodent bioassays of glyphosate were likely to be false positives. In particular, these Panelists believe that the lack of corroborating information from bioassays using the same or higher doses and in the same rodent species, strain, and sex is important evidence against a genuine carcinogenic effect of glyphosate.

Other Panel members, however, believed that differences in study designs could explain some of the tumor response discrepancies, and that, overall, the rodent bioassay data were consistent with glyphosate acting as a weak tumor promoter. As mentioned in response to question 3b, a pooled analysis that appropriately addresses differences in study design is recommended by some Panel members to determine whether or not the evidence from multiple rodent studies is indeed conflicting.

There has been no direct test of the weak promoter hypothesis (such as in a standard initiation-promotion bioassay), and some Panel members felt that such a conclusion was

speculative. Nonetheless, some of the high-dose responses in some of the fifteen or more bioassays of glyphosate were noteworthy, at least to some Panelists. EPA mentioned that dose-rates of greater than 1,000 mg/kg-day are considered to be above the "limit dose," but, in the case of glyphosate, its remarkable non-toxicity means that it can be reliably applied in bioassays at up to several grams/kg-day. Thus, in these Panelists' opinions, positive results at doses at and above the "limit dose" should not necessarily be discounted. Other Panelists also believe that the limit dose is used as a study design guideline and should not be used to discount findings after a study has been conducted. See further discussion of the appropriateness of the limit dose criterion in response to charge question 3f.

Some Panelists noted that the Issue Paper correctly noted the lack of reproducibility of statistically significant carcinogenic effects in the glyphosate rodent bioassays. However, this was noted only in a single sentence at the end of the discussions of the evidence from rat and mouse data (Section 4.8). These Panelists suggest that it would be helpful to substantiate this point by reviewing the information that led to this conclusion. To this end, these Panelists provide, in the following paragraphs, their summaries of the data from the rodent glyphosate studies available to the Panel on each of the cancer endpoints analyzed in the Issue Paper.

In Table 4.1 of the Issue Paper, the response of testicular interstitial cell tumors (also known as Leydig cell tumors) in male Sprague-Dawley rats (Lankas, 1981) had a significant dose response trend ($P = 0.0062$ by the exact form of Cochran-Armitage trend test (Gart et al. 1986, p. 81-86)). The highest dose in Lankas et al. was only 31 mg/kg/day, and for this reason was given the lowest rating (Klimisch 3) by Greim et al. (2015). Other studies of Sprague-Dawley rats using doses 30-fold higher or greater, from 940 mg/kg/day to in excess of 1000 mg/kg/day (Stout and Ruecker 1990; Atkinson 1983a; Enemoto 1997), did not show statistical evidence of an effect on testicular interstitial cell tumors. Indeed, in all of the 15 rodent bioassays, increases in Leydig cell tumors were seen only once.

In Table 4.2 of the Issue Paper, the response of pancreatic islet cell adenoma in male Sprague-Dawley rats (Stout and Ruecker 1990, high dose ≈ 1000 mg/kg/day) was analyzed. Although the trend test was not significant ($P = 0.18$) this analysis was included because of a significant increase at the lowest dose in a pair-wise test before adjusting for multiple comparisons ($P = 0.018$). The trend test for the combined response of adenoma or carcinoma gave a $P = 0.21$, not significant. However, in this same study there was a significant negative trend in female mice for both adenoma ($P = 0.04$, negative trend) and adenoma or carcinoma combined ($P = 0.04$, negative trend). Thus, overall, there appears to be greater evidence for a protective effect on pancreatic islet cell tumors in this study than a carcinogenic effect. However, these results most likely stem from natural variation in tumor responses, rather than any effect of treatment. Atkinson (1983a, high dose ≈ 1150 mg/kg/day) found a significant negative trend for this tumor among male Sprague-Dawley rats ($P = 0.007$, negative trend) using doses spanning those in Stout and Ruecker (1990). Enemoto (1997, reported in Greim 2015, high dose ≈ 1130 mg/kg/day) exposed male and female Sprague-Dawley rats to doses that spanned those used by Stout and Ruecker (1990) and found no statistical evidence of a dose effect. Also, two studies of male and female Wistar rats found no statistical evidence of a dose effect on pancreatic islet cell adenoma or adenoma and carcinoma (Brammer 2001, high dose ≈1300 mg/kg/day; Wood et al. 2009a, high dose ≈ 1100 mg/kg/day).

In Table 4.4 of the Issue Paper, hepatocellular adenoma in male Sprague-Dawley rats (Stout and Ruecker 1990, high dose ≈ 1000 mg/kg/day) was shown to have a significant dose response trend (P = 0.02), and, as a result, the combined response of hepatocellular adenoma and carcinoma was nearly significant (P = 0.078). In this same study there was an almost significant negative trend in hepatocellular adenoma among females (P = 0.078, negative trend). Atkinson et al. (1983a, high dose = 1000 mg/kg/day) and Enemoto (1997, as reported in Greim 2015, high dose ≈ 1200 mg/kg/day) also exposed Sprague-Dawley rats to doses that spanned those used by Stout and Ruecker (1990) and found no statistical evidence of a positive dose effect on adenoma or the combination of adenoma or carcinoma. In the Issue Paper, Table 4.9, hepatocellular adenoma in male Wistar rats (Brammer 2001, high dose ≈ 1300 mg/kg/day) had a significant dose response trend (P = 0.0082). However, Suresh (1996, high dose = 740 mg/kg/day) and Wood et al. (2009a, high dose ≈ 1300 mg/kg/day) exposed Wistar rats and found no statistical evidence of a positive dose effect, although there was an almost significant negative trend in hepatocellular adenoma for females in Suresh (1996) (P = 0.08, negative trend).

In Table 4.7 of the Issue Paper, thyroid C-cell adenoma in female Sprague-Dawley rats (Stout and Ruecker, 1990, high dose ≈ 1000 mg/kg/day) had a significant dose response trend (P = 0.040) and the trend in the combined response of adenoma or carcinoma was also significant (P = 0.042). The corresponding trends in male Sprague-Dawley rats were almost significant (adenoma response P = 0.079 and combined response P = 0.087). Atkinson et al. (1983a, high dose ≈ 1000 mg/kg/day) also exposed Sprague-Dawley male and female rats to comparable high doses with no statistical evidence of a dose effect on these tumors. There was a statistically significant trend (P = 0.0026) in thyroid C-cell carcinoma among female Sprague-Dawley rats in Lankas et al. (1981, high dose = 34 mg/kg/day), however the high dose in this study was well below the low dose in Stout and Ruecker (1990). Suresh (1996, high dose ≈ 740 mg/kg/day), Brammer (2001, high dose ≈ 1200 mg/kg/day) and Wood et al. (2009a, high dose ≈ 1200 mg/kg/day) exposed Wistar rats and found no statistical evidence of a positive dose effect on C-cell tumors. However, in Wood et al. among female Wistar rats there was a statistically significant negative trend in thyroid C-cell adenoma (recorded as parafollicular adenoma (Greim et al. 2015)) (P = 0.0030, negative trend), and a statistically significant negative trend in the combined response of C-cell adenoma or C-cell carcinoma (recorded as parafollicular carcinoma) (P = 0.0021, negative trend). Similarly, among male Wistar rats in Wood et al. (2009a) there were almost significant negative trends for thyroid C-cell carcinoma (P = 0.062, negative trend) and the combined response of thyroid C-cell adenoma or carcinoma (P = 0.064, negative trend).

In Table 4.10 of the Issue Paper, mammary adenocarcinoma in female Sprague-Dawley rats (Wood et al. 2009a, high dose ≈ 1200 mg/kg/day) had a significant dose response trend (P = 0.042) and the trend in the combined response of adenoma or adenocarcinoma was also significant (P = 0.0067). Brammer (2001, high dose = 1500 mg/kg/day) and Suresh (1996, high dose = 740 mg/kg/day) also exposed female Sprague-Dawley rats and found no statistical evidence of a positive effect on mammary tumors. However, there was a significant negative trend in mammary gland adenocarcinoma among female Sprague-Dawley rats in Suresh (1996 high dose = 741 mg/kg/day) (P = 0.018, negative trend).

In Table 4.12 of the Issue Paper, adenoma or carcinoma in tubule cell tumors in male CD-1 mice (Knezevich and Hogan, 1983, high dose ≈ 5000 mg/kg/day) had a non-significant dose-

response trend (P = 0.065). None of these tumors were seen in females. Atkinson et al. (1993a, high dose ≈ 1000 mg/kg/day) found two tubule cell adenomas in CD-1 male mice which occurred in control and low dose groups. Wood et al. (2009b, high dose ≈ 900 mg/kg/day) also exposed CD-1 mice, but did not record any incidences of tubule cell tumors.

In Table 4.14 of the Issue Paper, haemangiosarcoma in male CD-1 mice (Atkinson, et al., 1993b, high dose = 1000 mg/kg/day) showed a significant dose-response trend (P = 0.002), resulting from a total of four haemangiosarcomas, all appearing in the high dose group. Three haemangiosarcomas were detected in female mice but their distribution in the dose groups did not suggest a dose-response trend. Knezevich and Hogan (1983, high dose ≈ 5000 mg/kg/day) did not report finding any haemangiosarcomas among male or female CD-1 mice. Wood et al. (2009b, high dose ≈ 900 mg/kg/day) reported finding 6 haemangiosarcomas in male CD-1 mice (2 in controls, 1 at the low dose, 2 at the middle dose and 1 at the high dose) and four haemangiosarcomas in female mice (1 in each dose group).

In Table 4.15 of the Issue Paper, the incidences of lung tumors are recorded for male CD-1 mice (Wood et al. 2009b, high dose ≈ 900 mg/kg/day). A significant trend is reported for adenocarcinoma (P = 0.028), but not for adenoma or the combination of adenoma and adenocarcinoma. There is an almost significant negative trend for adenoma (P =0.078, negative trend). Knezevich and Hogan (1983, high dose ≈ 5000 mg/kg/day) found a highly significant negative trend for adenoma in female CD-1 mice (P = 0.00058, negative trend) and a significant negative trend for the combination of adenoma and adenocarcinoma (P = 0.015, negative trend). They also found a near significant negative trend for adenocarcinoma in male mice (P = 0.094, negative trend). Neither Atkinson et al. (1993b, high dose = 1000 mg/kg/day) nor Sugimoto (1997, high dose ≈ 4200 mg/kg/day) found statistical evidence for an effect on lung tumors in either male or female CD-1 mice.

In Table 4.16, of the Issue Paper, malignant lymphomas in male CD-1 mice (Wood et al., 2009b, high dose ≈ 900 mg/kg/day) had a significant dose response trend (P = 0.0066). Sugimoto (1997, high dose ≈ 4200 mg/kg/day) also found a significant trend (P = 0.016) for malignant lymphoma in male CD-1 mice. However, the two responses do not appear congruent: Wood et al. found a significant excess (5/51) at a dose of 810 mg/kg/day, whereas Sugimoto found no tumors (0/50) at a comparable dose (838 mg/kg/day). Knezevich and Hogan (1983, high dose ≈ 5000 mg/kg/day) found an almost significant trend in malignant lymphoma (P = 0.063) among female CD-1 mice, but there was no evidence of a positive trend in males (equal responses of 2 animals in both the highest dose group (4841 mg/kg/day) and in controls).

In Table 4.17 of the Issue Paper, haemangiomas in female CD-1 mice (Sugimoto 1997, high dose ≈ 4100 mg/kg/day) showed a significant dose-response trend (P = 0.0022). (This study was not available to the Panel.) Neither Knezevich and Hogan (1983, high dose ≈ 5000 mg/kg/day) nor Wood et al. (2009b, high dose ≈ 900 mg/kg/day) found any statistical evidence for an effect of treatment on the incidence of haemangiomas in either CD-1 male or female mice.

The Panel observed that an explanation needs to be provided in the Issue Paper of the criteria used to select tumors for detailed evaluation. Was a well-defined tumor selection procedure used (such as a minimum number of total tumors), or was it based on the original authors' evaluations?

Following this assessment, some Panelists concluded that the Issue Paper correctly finds the tumor-response data to be too inconsistent to be credibly considered to be compound-related. In many cases there were significant or near significant negative trends in the same tumor categories as those in which significant positive trends were identified. As noted elsewhere, with so many tumor categories recorded in these studies, a few significantly positive trends (and significantly negative trends) would be expected in each study even if the treatment has no effect on tumor rates. This multiple comparison problem is particularly acute in the case of glyphosate, because of the exceptionally large number of rodent studies available. This conclusion is consistent with the assessment presented by Dr. Haseman before the Panel, which showed the number of significant trends (at both significance levels of 0.05 and 0.01) in the animal study data are no greater than what would be expected simply due to the random assignment of animals to dose groups without any carcinogenic effect of glyphosate exposure.

Other Panelists pointed out that the Issue Paper ascribes equal weight to the 15 acceptable rodent carcinogenicity studies and concludes that "tumors seen in individual [rat or mouse] studies were not reproduced in other studies conducted in the same animal species and strain at similar or higher doses." These Panelists noted that in order to judge whether this conclusion is valid (and should be given more weight than the positive tumor findings), one must consider whether the studies were of similar quality, employed rodents with equivalent tumor sensitivities, and whether equivalent tumor incidence data was analyzed in a consistent manner. The Panel collected data from the primary study documents that suggests the studies varied with respect to these criteria. Some observations about how these studies varied in meaningful ways that could affect direct comparability, such as the direct comparisons made above, are provided in the following paragraphs.

The studies varied in terms of design and quality, in ways expected to impact their sensitivity. For example, the study by Lankas (1981) [MRID 00093879] treated rats for 26 months, which may explain why it detected a tumor response not detected in the other studies that treated rats for 24 months. The Stout and Ruecker (1990) rat study generated statistically significant responses for three different tumor types. This study may have had greater sensitivity than others because it employed 60 rats/treatment group (the largest number of rodents in any glyphosate study). In this regard, it should be noted that the glyphosate Issue Paper statement "…tumors at multiple sites…" is an observation that adds strength to the significance of tumor findings in carcinogenicity studies.

Across the mouse studies, mice were exposed through the diet for between 16 to 24 months. The mouse study by Reyna and Gordon (1973) sacrificed males after 16 months, females after 18 months and included histopathological analyses on only 10 mice per dose. Clearly, this study should not be weighted as heavily as studies where histopathology results were obtained from all 50 animals/sex/dose.

Some of the studies had low survival at terminal euthanizing (<20 animals/group), which is expected to reduce the sensitivity of the bioassay.

The study by Pavkov and Wyand (1987) [MRIDs: 40214007, 41209905, and 41209907] used a distinct test article and vehicle (Sulfonate, glyphosate trimesium salt and a 1% propylene glycol vehicle). The mouse bioassay by Pavkov and Turnier (1987) [MRIDs: 40214006 and

41209907] also employed Sulfonate as the test article and propylene glycol as the vehicle. It is inappropriate to consider a study as a reproduction of another study if a different test article was used.

In the Pavkov and Turnier study, males in the 0 ppm treatment group were euthanized after 89 weeks of treatment, whereas mice in the other treatment groups were euthanized after 95 weeks of treatment.

It was not clear to the Panel how tumor responses were systematically examined and reported by research pathologists across studies. For the majority of the studies presented in the Issue Paper, the combined incidence of euthanized in extremis/found dead plus terminally-euthanized animals were used in the analyses. In some cases, the Issue Paper assessment excluded animals that died before 55 weeks, but in other cases the use exclusion/inclusion status of these animals is unknown. For example, the Pavkov and Wyand and Pavkov and Turnier studies combined data on interim euthanized (6, 12, and 18 months) with moribund/dead and euthanized post experimental termination for statistical analysis. For Stout and Ruecker, the data are broken out as scheduled euthanizing (12 and 24 months?), unscheduled deaths, and "all deaths reported." The Panel found it was unclear from reading the Issue Paper whether analysts were able to parse out and analyze equivalent datasets using the same statistical approach for all 15 rodent carcinogenicity datasets. The Panel recommends the Issue Paper adds a section with a detailed description of which data were extracted and precisely how data were selected for subsequent analysis.

The Panel found no discussion in the Issue Paper of the extent to which histopathological examinations were performed in an equivalent manner across studies. The rat bioassay by Suresh (1996) [MRID 49987401] did not include histopathological analyses on all the low and mid-dose rats at terminal euthanizing and reported that "autolysis precludes evaluation" of many samples.

The Panel noted that it is also important to consider genetic variability across rodent strains used and how this variability impacts bioassay reproducibility.

Rodent strains maintained as separate breeding colonies for extended periods of time do not necessarily have the same spontaneous tumor profiles (King-Herbert and Thayer, 2006). This is the basis for the OECD recommendation that only studies performed within five years in the same laboratory should be considered as historical controls. To evaluate the variability among the rodents used, the incidence of a single tumor type in control rodents was compared across glyphosate studies.

- Male Sprague-Dawley rats, pituitary tumor rates reported as 40%, 56.6%, 58%, 70% and 52%.

- Female Sprague-Dawley rats, pituitary tumors rates reported as 88%, 76.7%, 81.6%, 94%, and 72%.

- Male Wistar rats, pituitary tumor rates reported as 30%, 34%, and 6%.

- Female Wistar rats, pituitary tumors rates reported as 80%, 47%, and 16%.

- Male CD-1 mice, pituitary tumors rates reported as 64%, 0%, 0%, and 0%.

- Female CD-1 mice, pituitary tumors rates reported as 64%, 2%, 0%, and 0%.

These data suggest that even within a particular rodent species there are relatively large differences in background tumor incidence rates which might be expected to impact the detection of statistically-significant findings.

Reported toxicological findings also varied across the different tumor bioassays, providing additional evidence of biological and/or methodological variability in the studies conducted in the US, the UK, Japan, and India between 1973 and 2009.

Most importantly, before one can conclude that the findings in individual studies are not replicated, one must compare the results across studies in a rigorous manner. Similar patterns of tumor responses were observed across studies for some tumor types.

- Lung: six studies in which all glyphosate treated groups have an equal or greater tumor incidence than the concurrent control group (for at least one tumor type in one sex), with the highest observed tumor incidence approximately twice the control level.

- Liver: five studies in which all glyphosate treated groups have an equal or greater tumor incidence than the respective control group (for at least one type of tumor in one sex) and the highest observed tumor incidence is approximately twice the control level.

- Lymphatic and thyroid tumors: three studies in which all glyphosate treated groups have an equal or greater tumor incidence than the respective control group (for at least one type of tumor in one sex) and the highest observed tumor incidence is approximately twice the control level.

One Panel member was of the opinion that this constitutes reproducible evidence of a biologically-significant carcinogenic effect in rodent liver, lung, thyroid, and lymphoid cells.

f. As described in Section 1.4, high-end estimates of exposure based on the currently registered uses for glyphosate in the United States have been calculated as 0.47 mg/kg/day and 7 mg/kg/day for potential residential and occupational exposures, respectively. As a result, the agency concluded that tumors observed at high-doses (approaching or exceeding 1,000 mg/kg/day) following glyphosate administration are not relevant for human health risk assessment. Please comment on the conclusions regarding the relevance of high-dose tumors to the human health risk assessment for glyphosate.

Panel Response

The EPA (2016a) evaluation defined 1,000 mg/kg/day as the “limit dose” and high-dose tumors (e.g., tumors in animals exposed to greater than 1,000 mg/kg/day) were given less weight. All but five of the 15 rodent studies investigated involved doses that exceeded 1,000

mg/kg/day. However, the EPA (2005) *Cancer Guidelines* suggest that an excessively high dose would be "5% of the test substance in the feed for dietary studies." None of the 15 studies utilized a dose as high as 5%. The highest tested dose in any of the 15 studies appears to be 30,000 ppm (3%) in the Knezevich and Hogan (1983) CD-1 mouse study. Therefore, at least based on EPA (2005) *Cancer Guidelines*, some members of the Panel concluded it is questionable whether results from exposures greater than 1,000 mg/kg/day, but less than doses corresponding to 5% in diet, should be given less weight. Disregarding responses at any dose above a pre-selected "limit dose," even though the dose did not exceed the maximum tolerated dose (MTD), is not in keeping with the way rodent bioassays are normally interpreted, which is to answer the question "was the test material carcinogenic in this study (assuming that the study did not use doses that exceeded the MTD)." Thus selecting 1,000 mg/kg/day *a priori* as the limit dose appears to be an *ad hoc* decision that is not well-justified, and is not justified on the basis of the EPA (2005) *Cancer Guidelines*.

One Panel member noted the possibility whereby a carcinogenic response at 1,000 mg/kg/day could, depending on the dose-response shape at lower doses, translate into an estimated human risk as high as 1% from a lifetime exposure to 7 mg/kg/day, the stated maximum potential exposure level for occupational exposure (assuming a 10% risk at 1,000 mg/kg/day in a mouse bioassay, a linear dose-response to lower doses, and using a surface area conversion from mice to humans, which entails multiplying by roughly a factor of 13). Further, a carcinogenic effect at a dose >1,000 mg/kg/day could suggest a carcinogenic effect is also occurring at lower doses, which cannot be detected due to lack of power.

The Panel concluded that the Issue Paper needs to clarify its position on results from exposures that exceed 1,000 mg/kg/day. In at least one place the document says that the tumor responses, including those from doses exceeding 1,000 mg/kg/day, are not related to treatment. For example, "in 5 of the 9 rat studies conducted with glyphosate, no tumors were identified for detailed evaluation. Of the remaining 4 rat studies, a statistically significant trend was observed for tumor incidences in the testes, pancreas, liver, thyroid, or mammary gland; however, the agency determined that these tumor findings are not considered to be related to treatment…." The statistically significant trends discussed in this paragraph are based on data at all the doses, including those exceeding 1,000 mg/kg/day. Thus, this paragraph indicates that the Issue Paper does not consider the tumor responses at any dose, including those exceeding 1,000 mg/kg/day, to be related to treatment. On the other hand, there are numerous statements in the report that suggest that tumors occurring at doses exceeding 1,000 mg/kg/day are related to dose:

"…tumor incidence in animal carcinogenicity studies was typically only increased at the highest doses tested (≥1,000 mg/kg/day)."

"… however, the data are not sufficient to determine whether linear kinetics is occurring at high doses where tumors were observed in animal carcinogenicity studies."

"Tumor incidences were not increased in animal carcinogenicity at doses <500 mg/kg/day, …"

"In the remaining studies, tumor incidences were not increased at doses <500 mg/kg/day, except for the testicular tumors observed in a single study. Increased tumor incidences at or exceeding the limit dose (≥1000 mg/kg/day) are not considered."

"even though tumors were observed in animal carcinogenicity studies, the possibility of being exposed to these excessive dietary doses …"

The Panel found that these statements suggest that the Issue Paper considers some tumors occurring at >1,000 mg/kg/day to be related to dose.

Many on the Panel expressed concern that not considering tumor responses at doses exceeding 1,000/mg/kg/day is not consistent with either EPA (2005) *Cancer Guidelines* or standard ways in which bioassay results are typically interpreted. However, the Panel also noted that tumors induced at only very high doses are less of a safety concern than those induced at doses within the range of human exposure; though one Panel member noted that it is very likely that workers in manufacturing/formulation and wholesale handling and also persons involved in accidents and spills may experience these high exposures.

There were some differences of opinion among Panel members regarding the relevance and use of high-dose tumors for human health risk assessment of glyphosate. Some Panel members felt that at high doses homeostatic mechanisms can be overwhelmed (e.g., the saturation of elimination processes) and, hence, allowances need to be made in the interpretation of high dose data. Other Panelists felt that data from high doses should be included with a caveat that high doses could lead to toxicity, but not carcinogenicity. Still other Panelists felt that if there were tumors in the presence of other toxicity at high doses, that would be cause for concern regarding the interpretation of the results and potentially justification for excluding the data.

One Panel member noted that effects consistent with carcinogenic potential occurred at doses lower than 1,000 mg/kg/day. Specifically, significant induction of lymphocytic hyperplasia was observed at 11 mg/kg/day (Lankas, 1981). Significant lymphoid hyperplasia was observed at low and mid doses in males (71.4 and 234.2 mg/kg/day) in a study where malignant lymphomas were significantly induced at 810 mg/kg/day (Wood, 2009a). Male Sprague-Dawley rats in the Lankas study (1981) demonstrated a significant trend and a significant pairwise comparison between control and the high dose for testicular interstitial tumors, when the high dose was 31.49 mg/kg/day. A significant pairwise comparison relative to controls was observed for pancreatic islet cell tumors in male Sprague-Dawley rats at the low dose, 89 mg/kg-bw/day (Stout and Ruecker, 1990). In the view of one Panel member, these carcinogenic effects in rodents should be considered when setting acceptable levels of glyphosate exposure.

As a general matter, EPA usually does not consider tumors observed at high doses in rodent bioassays to be necessarily irrelevant for purposes of human cancer risk assessment. However, if these observations are in fact false positives, then discounting them would indeed be appropriate. Making a *post hoc* assessment to discount high doses after the studies were designed and carried out using existing guidelines poses a concern.

The Panel noted that one bioassay did report a positive result (for Leydig cell tumors in a group of Sprague-Dawley rats) at a glyphosate-dose of only 31 mg/kg/day. However, some Panel

members noted that Leydig cell tumors (i) are quite common in Sprague-Dawley rats, (ii) are very difficult to distinguish from simple hyperplasia, and (iii) were not found to be associated with glyphosate in any of the other 14 bioassays, including those that used the same species, strain, and sex -- even at doses at and above 1,000 mg/kg-day.

One Panel member also noted that glyphosate acid (the form used in the bioassays) has a pH of 2 at saturation, so that very high-dose responses might be due to simple acidity, rather than to a compound-specific effect. Since glyphosate as used in commerce is not an acid but instead a more neutral salt or zwitterion, such effects might well be irrelevant for purposes of human cancer risk assessment.

g. Please comment on the strengths and uncertainties associated with the agency's overall weight-of-evidence and conclusions based on the available animal carcinogenicity studies, as described in Section 4.8.

Panel Response

Some Panelists felt that the Issue Paper did a good job discussing strengths and uncertainties of the animal carcinogenicity studies whereas others disagreed with the interpretation of the rodent carcinogenicity data as presented in EPA's Issue Paper. Responses to the previous sub-questions of Charge Question 3 give detailed discussion of the Panel's view of the considerations that went into the weight-of-evidence analysis.

The Panel members who disagreed with the Agency's interpretation of the rodent carcinogenicity data felt that the EPA (2016a) weight-of-evidence evaluation gave excessive weight to several factors: monotonic dose responses, historical tumor rates, lack of statistical significance in pair-wise comparisons when there is a significant trend, and disregarding or giving low weight to results at exposures > 1000 mg/kg/day.

Most Panelists were in general agreement that EPA's weight-of-evidence did not properly address the unusually large number of bioassays available for glyphosate. Some Panelists concluded that EPA ignored the serious multiple comparison problem caused by focusing attention on the most extreme tumor responses out of a large number of responses. Some Panelists determined that the best way to address this concern was to conduct properly pooled analyses, in order to determine the most conclusive answer to the question of whether there was evidence of a carcinogenic effect in any cancer endpoint, in any species or gender. See the extensive discussion of the Panel's thoughts on how to address the unusually large number of bioassays in response to Charge Question 3b.

Still other panelists, while agreeing that pooling can have the beneficial effect of allowing results in the same cancer endpoint from different sets of data to reinforce one another, pooling does not address the multiple comparison problem *per se*. In a large data set such as exists for glyphosate there will be a large number of responses that can be pooled, and consequently the multiple comparison problem will remain. This problem can be addressed by applying statistical tests specifically designed for tumor bioassay data that provide a single valid P-value for a tumorigenic effect at any site, which can also be designed to allow results in the same cancer endpoint from different sets of data to reinforce one another.

Each of the considerations for the weight-of-evidence analysis called out above are summarized briefly in the remainder of this response. See previous Charge Question responses for additional details.

Monotonicity

The Panel noted that the fact that an observed dose-response is not monotone typically provides essentially no evidence that the underlying true dose response is non-monotone. Furthermore, checking for monotonicity is not mentioned in EPA (2005) *Cancer Guidelines*. In the simulated examples reported earlier the probability of a non-monotone was 0.57 and 0.70 even though the true dose responses were monotone increasing or non-decreasing.

Historical Control Rates

In cases in which the Issue Paper incorporated historical control rates in setting weights, it was used to down-weight a significant tumor response as not dose-related. If this is true, then all the tumor responses observed in all groups are incidental. Thus, it would be reasonable, if the conclusion is that the tumor response is not dose-related, to compare the responses in all the animals in a study to historical controls, not just those assigned randomly to the study control group. The EPA (2005) *Cancer Guidelines* properly recommend caution in the use of historical control data: "Generally speaking, statistically significant increases in tumors should not be discounted simply because incidence rates in the treated groups are within the range of historical controls or because incidence rates in the concurrent controls are somewhat lower than average. Random assignment of animals to groups and proper statistical procedures provide assurance that statistically significant results are unlikely to be due to chance alone." Moreover, a careful review of historical control data to ensure comparability directed by the *Guidelines* was not described. Thus, some Panelists found that reliance on historical control data in the EPA (2016a) weight-of-evidence evaluation was overdone, and not done in accordance with EPA (2005) *Guidelines*.

Pairwise Tests

In several cases EPA (2016a) used the non-significance of pairwise tests to down-weight a significant trend test. This is contrary to EPA (2005) *Cancer Guidelines* which state: "Significance in either kind of test is sufficient to reject the hypothesis that chance accounts for the result." As noted above, the EPA (2016a) analysis would be on a sounder and more easily interpreted footing if it eschewed a battery of pairwise tests, and instead conducted a single powerful test for carcinogenicity, namely a trend test.

Disregard of exposures > 1000 mg/kg/day

The EPA (2016a) practice of disregarding or giving low weight to results at exposures > 1000 mg/kg/day, seems to be at odds with the EPA (2005) *Cancer Guidelines*, which suggest that an excessively high dose would be "5% of the test substance in the feed for dietary studies." But 5% in feed is considerably greater than 1000 mg/kg/day in both rats and mice, and none of the doses utilized in the studies reviewed exceeded 5% in feed. Several Panel members saw no overriding reason for disregarding results from exposures > 1000 mg/kg/day, so long as the dose does not exceed the maximum tolerated dose. These Panelists also thought that responses at such

doses may have relevance for human risk, as the presence of a response at a high dose suggests that there possibly could be risks at lower doses that were too low to be detected because of reduced power at lower doses. One Panelist commented that the definition of limit dose was for the design of animal bioassay studies, and should not be applied to the interpretation of the results of studies designed under the *Guidelines*.

Weight-of-evidence assessment in the presence of a large number of bioassays

All of the shortcomings discussed in the previous paragraphs are in the direction of making a conclusion of no carcinogenic effect. However, in the opinion of some Panel members, it seems likely that these shortcomings are more than compensated for by focusing on the statistical significance of the tumor types showing most extreme dose-responses among a very large number of tumor types for which data are available. With such a large number of tumor types available for statistical evaluation in an animal study (e.g., 100+), several tumor types would be expected to be significant at the $P = 0.05$ level, even if the treatment has no effect on tumor occurrence. Other Panelists recommended two other perspectives for determining weight-of-evidence in the presence of a large number of bioassays: one perspective is based on interpretation of the current wording of the *Guidelines*, and the other recommends the Agency implement pooled analyses. See the discussion in response to charge question 3b for further details describing three different approaches to handling multiple bioassays and the justification for these various approaches.

Concluding comments

The EPA Issue Paper concluded that the observed tumor responses correspond to common spontaneous tumor types, which are unrelated to glyphosate treatment. Some Panel members agreed. Other Panel members interpreted the totality of the tumor data as supporting the hypothesis that glyphosate causes the promotion or progression of common spontaneous lesions and considered the observed degree of bioassay irreproducibility is expected given differences in rodent genetics, study design and study quality. Some Panel members agreed that there is sufficient evidence to conclude glyphosate is a weak rodent carcinogen.

4. As part of its analysis, the agency has considered almost 200 assays investigating the genotoxic potential of glyphosate. Of these, 107 were performed with the active ingredient glyphosate. These included *in vitro* and *in vivo* studies from the open literature, as well as studies submitted to the agency that were conducted according to Office of Chemical Safety and Pollution Prevention (OCSPP)/ Organization for Economic Cooperation and Development (OECD) guidelines. Non-mammalian studies were excluded from this analysis unless the assays were generally recognized to inform the human carcinogenic potential of glyphosate (e.g., bacterial reverse mutation assays). Studies evaluated genotoxic endpoints, such as gene mutations in bacteria and mammalian cells, chromosomal aberrations, micronuclei formation, and other assays measuring DNA damage.

a. Please comment on the agency's review and evaluation process of relevant genotoxicity studies to inform the human carcinogenic potential of glyphosate, including the decision to exclude non-mammalian studies (e.g., reptiles, plants, worms, fish), except those generally recognized to inform human carcinogenic potential.

Panel Response

Panel members agreed that the review and evaluation process of genotoxicity studies is sufficient given the limits of the accepted assays, which are described in the report (first paragraph of section 5.1) as being sufficient to detect: "1) changes in single base pairs, partial, single or multiple genes, or chromosomes, 2) breaks in chromosomes that result in transmissible deletion, duplication or rearrangement of chromosome segments, and 3) mitotic recombination."

One Panel member recommended that this section of the document be expanded to indicate that none of the assays employed provides an unbiased (global) measure of small insertions, deletions and rearrangements, which can result in gene copy number variation (CNV). CNVs are best resolved using sequence-based approaches and are important for several reasons. CNVs are now known to occur at greater rates than other types of mutations and can arise both meiotically and somatically. CNVs arise via mechanisms that differ from base-substitution mutations, including inhibition of replication, which some studies have reported for glyphosate (Marc et al., 2002 and 2004). Structural mutations may contribute to human variation at least as much or more than base-substitution mutations. Further, strong associations have been observed between CNVs and many cancers, cancer risk factors, and mechanisms for promotion. Finally, there seems to be some evidence that structural mutations contribute to response to glyphosate exposure (Gaines et al., 2010; Widholm et al., 2001).

One Panel member encouraged the agency to consider two key human biomonitoring studies in their evaluation of genotoxicity, specifically studies by Bolognesi et al. (2009) and Koureas et al. (2014). Bolognesi et al. (2009) evaluated genotoxicity as binucleated micronuclei and observed some increases in the blood cells of Columbian farmers after aerial spraying of glyphosate. Koureas et al. (2014) measured oxidative DNA damage as 8-hydroxy-2' -deoxyguanosine (8-OH-dG) and reported that glyphosate applicators more often had high levels of 8-OH-dG than non-applicators (43.8% vs 27.9%, RR=1.47, 95% CI=0.78, 2.77).

Because some Panel members concluded that the rodent bioassay data indicates that at high dose, dietary exposure to glyphosate causes promotion/progression of pre-existing spontaneous lesions, studies in mammalian and non-mammalian species are of interest in terms of understanding potential underlying mechanisms of promotion/progression. Disruption of the proteome is one potential non-genotoxic mechanism of carcinogenesis. A recent study by Ford et al. (2017), concluded that *in vivo* glyphosate exposure may lead to generation of reactive metabolites such as glyoxylate which may in turn inhibit fatty acid oxidation enzymes, heighten levels of triglycerides and cholesteryl esters, and may potentially lead to metabolic disorders stemming from impaired fatty acid oxidation or fatty acid metabolism, including obesity, hepatic steatosis, atherosclerosis, and dyslipidemia.

Panel members agreed that, in the determination of whether or not glyphosate is likely to be genotoxic in humans, the EPA document focuses appropriately on studies conducted in cultured mammalian cells and exposed animal models.

b. Consistent with the OECD guidance (2015), *in vivo* findings in genetic toxicology testing are generally considered as having a greater relevance to humans than *in vitro* findings.

Consistent with the 2005 *Cancer Guidelines*, all available data were considered in the weight-of-evidence evaluation of the genotoxic potential for glyphosate. The relevant studies are summarized in Tables 5.1-5.7. Please comment on the agency's approach for evaluating the genotoxicity data.

Panel Response

The Panel found that the agency has assembled and evaluated relevant genotoxicity data in an appropriate manner, with the previously mentioned caveat regarding the lack of robust detection of CNVs.

c. As described in Section 1.4, oral exposure is considered the primary route of concern for glyphosate and high-end estimates of exposure range from 0.47-7 mg/kg/day. Please comment on the human health relevance of the genotoxicity findings with respect to the doses where effects were observed and the route of administration.

Panel Response

Panel members agreed that genotoxicity studies were conducted at sufficiently high doses (and range of doses). There is a sufficient number of negative studies, where glyphosate was administered through the oral route, to support the agency's conclusion that glyphosate is not genotoxic. A few positive findings in studies employing high dose exposures through the intraperitoneal (IP) route of administration may represent secondary effects of toxicity.

Several Panel members commented that if glyphosate causes progression of spontaneously arising lesions (cells carrying cancer driver mutations or other types of DNA damage), then humans are at risk of glyphosate-induced carcinogenicity and the longer human lifespan (as compared to rodents) is expected to contribute to the risk.

d. Please comment on the strengths and uncertainties associated with the agency's overall weight-of-evidence and conclusions based on the available genotoxicity studies, as described in Section 5.7.

Panel Response

Panel members found that the Agency's overall weight-of-evidence and conclusion that there is no convincing evidence that glyphosate induces mutations *in vivo* via the oral route are sound. Areas of remaining uncertainty are related to the potential for glyphosate-induced inflammation and genotoxic effects secondary to toxicity caused by high dose exposures (i.e., glyphosate-induced inflammation, oxidative stress, 8-OH-dG, and sister chromatid exchanges or SCE) and whether the glyphosate-containing formulations have genotoxic potential.

5. The modified Bradford Hill criteria were used to evaluate multiple lines of evidence using such concepts as strength, consistency, dose response, temporal concordance, and biological plausibility. In accordance with the 2005 *Cancer Guidelines*, the agency used a weight-of-evidence analysis to characterize the human carcinogenic potential of glyphosate and determine which cancer descriptor is supported by the data. The agency has described

the strengths and uncertainties associated with the choice of various cancer descriptors with a focus on "suggestive evidence of carcinogenic potential" and "not likely to be carcinogenic to humans". Please comment on the completeness, transparency, and scientific quality of the agency's characterization of the carcinogenic potential.

Panel Response

The Panel was asked to comment on the completeness, transparency and scientific quality of the argument presented in the EPA's Issue Paper leading to the conclusion (page 141) that "The strongest support is for 'not likely to be carcinogenic to humans' at doses relevant to human health risk assessment." The Issue Paper's goal is to describe the Agency's "comprehensive analysis of available data from submitted guideline studies and the open literature." (page 140)

The Panel noted that the conclusion on glyphosate carcinogenicity offered in the Issue Paper has two parts. The first part is a hazard statement; the second part is a risk characterization statement. Since the Issue Paper is not a full risk assessment of technical glyphosate as outlined in the 2005 *Guidelines for Carcinogen Risk Assessment*, the Issue Paper conclusion is best assessed as a hazard statement.

This Issue Paper is conceptually driven by the 2005 *Guidelines for Carcinogen Risk Assessment* which in turn incorporates the "modified Bradford Hill Criteria to evaluate strength, consistency, dose response, temporal concordance and biological plausibility of multiple lines of evidence in a weight-of-evidence analysis" (page 14). The Issue Paper also draws on the 2010 EPA OPP draft "*Framework for Incorporating Human Epidemiologic & Incident Data in Health Risk Assessment*" which also utilizes a modified Bradford Hill Criteria as applied specifically to epidemiologic data.

Completeness of Agency's carcinogenic potential characterization

For the epidemiology studies, the Agency followed its peer-reviewed guidelines on evaluation and use of epidemiology studies in risk assessment and reviewed:

- Study design, including study sample size and power to detect effects under consideration,
- The quality of the exposure assessment in epidemiology studies,
- The potential for differential and non-differential misclassification of effects or outcomes,
- The measurement and utilization (or not) of potential confounders,
- Potential biases and their impacts on observed associations, and
- The associated statistical analysis.

For the animal studies, the Issue Paper review followed standard practice and considered in their review, in order to ultimately select and interpret the findings from well-conducted, long-term animal studies:

- Study design, sample size, and adherence to quality guidelines,

- Statistical analysis and use of trend and multiple comparison testing protocols,

- Concurrence with historical control rates,

- Evidence of carcinogenicity through magnitude of tumor response, occurrence at multiple sites, in multiple strains or species, their progression, latency, and dose response, and,

- Absence of tumors.

For genotoxicity studies, the Issue Paper review also followed standard practice and considered:

- Test type and objective,

- Substance tested (e.g. technical glyphosate),

- Quality of the implementation of the study (adherence to standard study design, sample size, dose, use of positive and negative controls),

- Conditions under which the study was performed (solubility, pH, osmolarity, cytotoxicity, but also degree of blinding in evaluation of outcomes), and,

- Consistency among findings and support for particular MOA.

By any criteria, this list suggests a complete review. However, as the Panel notes in earlier sections of this report, there are aspects of EPA's approach and conclusions that it recommends altering. Missing are:

- Study data and results on workers engaged in manufacturing, formulating and handling and wholesale selling of glyphosate – although mentioned a number of times in the Panel's discussions, it is generally assumed that there are no worker epidemiology studies because none are reported in the Issue Paper.

- Other human incident data, such as reports on acute accidental exposures – the 2010 EPA OPP draft "*Framework for Incorporating Human Epidemiologic & Incident Data in Health Risk Assessment*" discusses the utility of other incident data. While incident data have little direct relevance to cancer outcomes, time trends suggesting increasing episodes of acute exposures can also be suggestive of increases in overall exposure over time, which can in turn affect inferences about the quality and biases in the human epidemiologic studies.

Transparency (interpreted by the Panel as honesty and openness) of Agency's carcinogenic potential characterization

The Panel did not have major issues in following the Agency's assessment. Section 6.6 of the Issue Paper is clear in laying out the Agency's argument for its final classification. Supplemental documents provided on the meeting docket allowed Panel members the information necessary to duplicate most analyses and verify most report claims if so desired. While this document indicates areas where Panel members disagree with the Agency's assessment (see next section), the Panel was able to identify documents and data from which Issue Paper claims originated. The Panel expressed concern that a barrier to full transparency exists in that some of the documents and associated data that are used by EPA in this assessment require special procedures for access and a few studies were not available to the Panel or the public. The Agency explained that FIFRA regulations are responsible for some of these limitations. Regardless, the Panel wondered whether the public could fully review and reproduce the conclusions reached by EPA.

Scientific quality of the Agency's carcinogenic potential characterization

Quality science is reproducible, free from distortion, credible, built on what is known (sound science), follows logical inferences, and is honest about what is achievable and the limits of available designs and data. While the Issue Paper does try to detail the design and data limitations of each study selected, some on the Panel believed that it does not provide sufficient details to support its conclusions (for example, see discussion of Charge Question 3e) and this negatively impacts the scientific quality of the report. In addition, many Panel members felt that some of the discussions of study design and data limitations provided in the Issue Paper introduced and used criteria that were not part of EPA *Guidelines* for these assessments, and this further reduces the credibility of the assessment.

The issue of distortion impacts the nature and quality of the inferences drawn in the Issue Paper. These inferences also depend upon a consistent application of the Bradford Hill Criteria, aspects of which are discussed in the next sections.

Dose-response and temporal concordance

With regard to the epidemiology data, the Issue Paper concludes: "Based on the weight-of-evidence, the agency cannot exclude chance and/or bias as an explanation for observed associations in the database. Due to study limitations and contradictory results across studies of at least equal quality, a conclusion regarding the association between glyphosate exposure and risk of NHL cannot be determined based on the available data" [page 68]. The 2005 *Guidelines* state (page A-2): "When cancer effects are not found in an exposed human population, this information by itself is not generally sufficient to conclude that the agent poses no carcinogenic hazard to this or other populations of potentially exposed humans, including susceptible subpopulations or life stages."

The Panel debated the Issue Paper's assignment of weights and arguments that lead to its conclusion of "no observed association". The issues considered by the Panel and on which there was disagreement are as follows.

1. Some Panel members noted that weight placed on the findings from the prospective Agricultural Health (cohort) Study (AHS, 2005) is too high given that this study has had only

limited follow-up to date, is subject to selection bias due to its approach to recruitment (i.e. preferentially excluding cases that occur before the enrollment in the cohort) and uses biased exposure quantification in the analysis. An updated report is expected to clarify some of these issues, and is expected within the year.

2. The Panel could not agree on the appropriate weights to be applied to the Eriksson et al. (2008) and McDuffie et al. (2001) studies. Some Panel Members felt that that these studies provide positive evidence of a dose response to glyphosate exposure and hence should be afforded greater weight. Other Panel Members noted that the potential impacts of recall and selection biases on study findings are not properly accounted for in the Issue Paper's assessment and as a result these studies are afforded too much weight.

3. Some Panel members noted that the Issue Paper's discussion of the change-in-use patterns and their impact on risk estimates expected from the epidemiology (pages 129-130) does not reflect a good understanding of the estimates being reported. The Issue Paper's analysis appears to be aligning assumptions about absolute production changes over time with epidemiological inferences for relative impacts of exposure in ways that many on the Panel felt cannot be supported. It makes assumptions about the change of patterns of exposure based on trends in sales (production trends), that do not necessarily reflect the more relevant (but unknown) pounds per worker per year (relevant exposure metric). Trends in production can only be speculated to affect usage at the worker level. In documents presented to the Panel, the Center for Food Safety provided a compilation from USDA data (EPA-HQ-OPP-2016-0385-0438) showing evidence, based on pounds per acre per year, that glyphosate use in the 1980's may have resulted in potentially higher per worker exposure than in the 1990's, including after the large increase in glyphosate production. This argument is in direct contradiction to the arguments made in the Issue Paper. The Issue Paper's arguments about exposure trends also ignore latency. Since all the arguments in the paragraph beginning at the bottom of p. 129 (and also made in Section 3.6) rely on speculation and unverified assumptions, the Panel recommends they be removed from the Issue Paper.

4. Some Panel members interpreted the meta-analysis results for NHL to be indicative of a suggestive effect of glyphosate. Other Panel members felt that the limitations of the retrospective case-control studies, particularly with respect to recall bias, selection bias, and confounding by other aspects of living or working on a farm, combined to render the meta-statistics unreliable.

For the rodent carcinogenicity assay data, the Issue Paper concludes (page 96), "based on the weight-of-evidence, the agency has determined that any tumor findings observed in the rat and mouse carcinogenicity studies for glyphosate are not considered treatment-related. Tumor findings observed at the highest doses tested were also not reproduced in studies in the same animal strain at similar or higher doses." The 2005 *Guidelines* state (page A-4): "When cancer effects are not found in well-conducted animal cancer studies in two or more appropriate species and other information does not support the carcinogenic potential of the agent, these data provide a basis for concluding that the agent is not likely to possess human carcinogenic potential, in the absence of human data to the contrary." The 2005 *Guidelines* also state (page A-3): "The default option is that positive effects in animal cancer studies indicate that the agent under study can have carcinogenic potential in humans."

The Panel debated at length the Issue Paper's interpretation of positive tumor responses in rodent bioassays. Not all Panel members agreed with the Issue Paper's conclusion that findings in rodent bioassays are not treatment-related. The discussion centered on which analysis results constitute a significant finding and which support a false positive finding. Not everyone agreed with how the Issue Paper assigned weights/influence to non-monotonic dose-responses, to significant pairwise comparisons, and to comparisons with historical control data. Some Panelists felt that the Issue Paper did not take into consideration the number of statistical tests performed when assessing tumor incidence across rodent bioassays. Also discussed was the inference possible from using pooled analyses to draw gender- and species-specific conclusions from the multiple studies on each endpoint, and the impact of using exact *P-values* versus approximate or asymptotic *P-values* in the standard test used to evaluate evidence for dose response. Finally, there was discussion on the appropriateness and desirability of down-weighting results at doses that exceeded 1000 mg/kg-BW/day.

Following these discussions, some Panel members interpreted the positive animal bioassay findings as likely false positives, resulting from the very large number of statistical tests conducted, and concluded that there is no credible evidence of a dose response, thus agreeing with the conclusions of the Issue Paper. Other Panel members interpreted the significant trend test results as well as other rodent bioassay findings as sufficient evidence to conclude that glyphosate is a rodent carcinogen. Under the 2005 *Guidelines* these findings are adequate to support a conclusion that glyphosate can have carcinogenic potential in humans (more details are provided in the section on *Evaluation and Proposed Conclusion* discussion below).

For the genotoxicity studies, the Issue Paper concludes (page 128): "The overall weight-of-evidence indicates that there is no convincing evidence that glyphosate induces mutations *in vivo* via the oral route" and "While there is limited evidence (of) genotoxic(ity) for effects in some *in vitro* experiments, *in vivo* effects were given more weight than *in vitro* effects particularly when the same genetic endpoint was measured, which is consistent with current OECD guidance. The only positive findings reported *in vivo* were seen at relatively high doses that are not relevant for human health risk assessment."

The Panel generally agreed with the Issue Paper's conclusion regarding the lack of genotoxicity effects of glyphosate.

Strength, consistency and specificity

With regard to the epidemiology studies, the Panel concurred with the Issue Paper conclusions regarding solid tumors, leukemia, MM and HL. Panel members differed in their agreement with the Issue Paper's conclusions regarding the strength, consistency and specificity of the epidemiological findings for a relationship between glyphosate exposure and NHL.

Some Panel members remarked on the consistency in the direction and value of the estimated odds ratios across multiple high-quality studies, illustrated by the very low heterogeneity statistic in the meta-analyses. They argued that the best quantification of the NHL evidence is provided by the meta-analyses with risk estimates of around 1.3 to 1.5 and with a lower bound around 1.0 but typically slightly above one when estimates are reported to 2 significant digits. Some Panel members found that the Issue Paper conclusion that the

epidemiology studies report conflicting results for NHL is based on a *post hoc* sorting of studies (see p. 130) that is not based on statistical evidence. The apparent inconsistency of the dose-response estimates is thus a result of this invalid and unsupported comparison, and this analysis finding should not be given any weight in the assessment. Some Panel members believed that the Issue Paper should rely on the meta-analysis results as the best estimate of the NHL effect, while other Panel members noted that meta-analyses cannot compensate for or remove residual confounding in the original studies. In other words, they stressed that if the individual studies present biased estimates, then the meta-result will also be biased.

The Panel discussion on the strength of findings in the animal carcinogenicity studies centered on the value of findings for doses that exceeded 1000 mg/kg-BW/day (the limit dose). The limit dose is used in the *Cancer Guidelines* as a design criterion (EPA 2005). Several Panelists noted that once an experiment is designed, interpretation of its findings should take into account all the data that were obtained under that design. The Panel consensus was that the Issue Paper needs to refine and strengthen its argument regarding this issue; it needs to clarify the use of the *Cancer Guidelines*, discuss in detail studies that might argue for or against the use of limit dose findings in tests, and ensure the 2005 *Cancer Guidelines* are adhered to in their revised approach. For many on the Panel, this criterion is viewed as having high potential to distort findings, and as such needs more careful discussion in the Issue Paper.

The Panel discussed at length the consistency, or lack thereof, of animal findings. In particular, some Panelists noted that although some individual rodent bioassays reported statistically significant results with regard to one or more tumor-types, these specific results were not replicated in other studies using the same rodent species and strain, even when replicate bioassays were run at higher doses. In evaluating the toxicology evidence from the rodent bioassays, pooling of animal data for dose-response modeling provides the equivalent of the meta-analysis performed for the epidemiology data. Dose response modeling results from pooled analysis, along with proper consideration of multiple comparisons, would provide a better base from which to discuss qualitatively the consistency of study findings for specific endpoints, species and in many cases, sexes.

One Panel member presented an argument that there is sufficient replication and magnitude of bioassay results to demonstrate treatment-related increases in rat hepatocellular adenoma, mouse lung adenoma/adenocarcinoma, and mouse lymphoma. The Panelist pointed to glyphosate-induced lymphoma in mice, where two of four studies (employing 50 mice/group) reported increases in lymphocytic hyperplasia in treated mice and three reported increases in lymphoma (including malignant lymphoma). Other Panel members agreed with this argument, particularly when examined with findings from a pooled analysis offered by a public commenter, Dr. Christopher Portier (EPA-HQ-OPP-2016-0385-0449).

Biological plausibility and coherence

Some Panel members recommended that the plausibility argument in the Issue Paper should be updated to address the hypothesis that glyphosate has potential to be a weak cancer promoter. Some on the Panel remarked that the hypothesis that glyphosate could be a weak promoter is not addressed in the Issue Paper. They feel that glyphosate as a weak promoter is a

potential explanation for the human and rodent study NHL and lymphoma results. These concerns also suggest the need for more discussion on immunotoxicity by glyphosate.

One Panelist commented that the epidemiology studies provide plausible evidence of a link between NHL occurrence and glyphosate exposure, a link that does not depend upon or require that the mechanisms driving this association are fully understood. This situation was compared to the evidence for air pollution health effects that are primarily based on findings from epidemiology studies, noting that it is only in recent years that mechanistic explanations have begun to emerge from toxicology and other experimental study findings to support these associations. In air pollution setting, relative risk estimates for an increment in pollution exposure are on the order of 1.03 for short-term acute effects estimated by time series studies, to 1.25 for longer-term chronic effects estimated in cohort studies. A similar evolution of understanding is plausible for glyphosate exposure.

Uncertainty

The Panel found that this section appropriately notes that the available database is remarkably large and should be adequate for evaluating carcinogenic potential but that many uncertainties remain. Some uncertainties brought forward from earlier sections in the Issue Paper, such as excluding formulations with glyphosate and weak pharmacokinetics could be expanded upon in this section. Uncertainties in epidemiological and animal study evidence are well discussed in earlier sections.

Some Panel members focused on the non-significance (i.e., *P-values* greater than 0.05) of odds ratios for the NHL analyses in the individual epidemiology studies, and that estimates of the lower bound of the 95 percent confidence interval of the odds ratio was at or below one across all studies. Various meta-risk estimates ranged between 1.3 and 1.5 and all had confidence interval lower bounds at or above 1.0. Given the potential for biases in estimates resulting from problems with exposure estimation, recall, and participant selection, as were pointed out in the Issue Paper, this relatively small but elevated risk with confidence interval lower bounds close to 1.0 did not argue for a strong and consistent finding of effect. In particular, in the opinion of some Panel members, the Issue Paper does not adequately assess the impact of potential biases on the odds ratio estimates, recall and selection bias in particular.

Evaluation and Proposed Conclusion

The Issue Paper concludes in Section 6.6.2 that the weight-of-evidence supports the descriptor for glyphosate of “not likely to be carcinogenic to humans” at the doses relevant to human health. The argument for concluding this classification rests on the assessment conclusion that there is “convincing evidence that carcinogenic effects are not likely below a defined dose range” where the data are “robust for deciding that there is no basis for human hazard concern.”

The Panel discussion mainly focused on how the Issue Paper did or did not argue for a hazard determination of “not likely to be carcinogenic to humans.” Most of the Panel’s discussion centered on assessment of the potential for glyphosate to be a carcinogen, and less on the conditions under which glyphosate exposure would represent a significant human health risk. In the Issue Paper, the statement “at doses relevant to human health” establishes a condition

under which glyphosate is "not likely to be carcinogenic to humans." The Panel discussions on the strength of association between glyphosate exposure and cancer incidence in the epidemiological studies, and the discussions on dose response in animal studies focused on assessing carcinogenic potential (at any dose) and less on establishing a threshold for risk to human health. This focus on the hazard identification is appropriate for an evaluation of the carcinogenic potential, as framed by EPA's 2005 *Guidelines* that describes hazard assessment as the first step of a risk assessment. For hazard evaluation, the question to be addressed is "Can the identified agent present a carcinogenic hazard to humans and, if so, under what circumstances?" (pp. 1-3)

The Panel was split between those members agreeing with the Issue Paper conclusions and those members who felt that the characterization of "not likely to be carcinogenic to humans" in the Issue Paper should be replaced by the hazard descriptor of "suggestive evidence of carcinogenic potential".

Perspectives supporting the "not likely to be carcinogenic to humans" descriptor

Some Panel Members concluded that while many of the issues identified in the Panel discussions can and should be addressed in the final EPA report on glyphosate, these changes would unlikely, in the opinion of these Panel members, change the final Issue Paper conclusions. They referenced a) a presentation before the panel by Dr. Haseman showing that the number of statistically significant responses in the glyphosate rodent bioassays is no greater than would be expected by chance, and b) a corroborating analysis presented by one Panel member. These arguments support their conclusion that the bioassay results are consistent with what would be expected by chance and not reflective of compound-induced effects. They see a wealth of studies with insufficiently consistent findings; several entirely negative bioassays, several weakly positive bioassay findings but not in the same tumor type, and, not in a majority of studies. These Panel members also concluded that the weakly positive results on NHL from the human case-control studies cannot be definitively linked to glyphosate-exposure, and biases from residual confounding due to other, non-glyphosate aspects of farming, recall bias, or selection bias are more likely explanations of these findings. The reproducible negative genotoxicity findings do not suggest a mutagenic MOA for glyphosate, and the Issue Paper presents no evidence that glyphosate is immunotoxic. Taken altogether, these findings are not sufficient to raise the hazard above "not likely to be carcinogenic to humans."

Some panel members did not agree with the premise that the rodent bioassay data indicate significant carcinogenic effects at doses that do not greatly exceed EPA's high-end estimate of occupational glyphosate exposure of 7 mg/kg-bw/day. In the opinion of these panel members, the rodent study results are more likely simply examples of the many incidental findings that are to be expected in a large database like the glyphosate animal database. Specifically:

1. In addition to the increases in the occurrence of lymphocytic hyperplasia cited in Lankas (1981) and referenced above there was a significant deficit of lymphatic hyperplasia among dosed animals at one site. More importantly, among Lankas (1981) and four other studies in Sprague-Dawley rats employing doses of between 30 and 40 times the highest dose used in Lankas (1981), none showed any evidence of an effect of glyphosate exposure upon lymphoid tumors.

2. In the lymphoid hyperplasia in Wood 2009b referenced above, lymphoid hyperplasia was detected in 4/36 male control mice, 1/1 low dose, 1/1 mid dose, and in 3/32 high dose male mice. Thus the claims of lymphoid hyperplasia at low and mid doses are both based on only one exposed animal each.

3. As noted in point 1, there are four bioassays in the same strain of rat employing doses of between 30 and 40 times the highest dose used in Lankas (1981). None of these bioassays show any evidence of an effect of glyphosate exposure upon testicular tumors (All of the responses in these bioassays are shown in Table 2).

4. The trend test was not significant in the response of pancreatic islet cell adenoma among male Sprague-Dawley rats in Stout and Ruecker 1990. However, significant negative trends occurred in female rats for both adenoma and adenoma and carcinoma combined. Two other studies in Sprague-Dawley rats and three in Wister rats all tested at a higher dose than Stout and Reucker, but found no evidence of a positive effect of glyphosate exposure upon these tumors. However, Atkinson 1983 found a highly significant negative trend in adenoma among male Sprague-Dawley rats.

Perspectives supporting the "suggestive evidence of carcinogenic potential" descriptor

Other Panel members did not agree with the conclusions of the Issue Paper. To these Panel members, the weight-of-evidence conclusion based on EPA's 2005 *Guidelines* naturally leads to suggestive evidence of potential carcinogenic effects. In their view, epidemiologic and rodent studies contain findings that together (coherence and consistency) suggest a potential for glyphosate to affect cancer incidence. Many of the arguments put forth in the Issue Paper discussion are not persuasive. These Panelists concluded that the epidemiologic and rodent study findings should not be discounted to the extent done in the Issue Paper. One Panel member argued that, using standard approaches in the analysis of the glyphosate rodent bioassay data, significant carcinogenic effects are observed at doses that do not greatly exceed EPA's high-end estimate of occupational exposure of 7 mg/kg-bw/day. Specifically noted are:

1. Lymphocytic hyperplasia at 11 mg/kg-bw/day in Sprague-Dawley rats in Lankas, 1981.

2. Lymphoid hyperplasia at low and mid doses in males at 71.4 and 234.2 mg/kg-bw/day in a study where malignant lymphomas were significantly induced at 810 mg/kg-bw/day in Wood et al. 2009b.

3. Testicular interstitial tumors in male Sprague-Dawley rats demonstrated a significant trend and a significant pairwise comparison between control and the high dose of 31.49 mg/kg-bw/day in Lankas, 1981.

4. Pancreatic islet cell adenoma in male Sprague-Dawley rats demonstrating a significant pairwise comparison relative to controls at the low dose, 89 mg/kg-bw/day in Stout and Ruecker, 1990.

According to the 2005 EPA *Guidelines for Carcinogen Risk Assessment*, the cancer descriptor "not likely to be carcinogenic to humans" applies if "there is convincing evidence that carcinogenic effects are not likely below a defined dose range." Many Panel members believe

that the EPA did not provide convincing evidence of a lack of carcinogenic effects. These Panelists agreed that the four findings listed above are adequate to reject the Issue Paper's conclusion of "not likely to be carcinogenic to humans" and support a conclusion of "suggestive evidence of carcinogenic potential" under these *Guidelines*.

Other perspectives

Some Panel members disagreed with the conclusion that the descriptor should be "suggestive evidence of carcinogenic potential" and some of these Panelists also did not feel comfortable with the descriptor "not likely to be carcinogenic to humans" either, preferring a descriptor such as "no credible evidence of carcinogenicity" or "equivocal."

REFERENCES

Acquavella, J; Garabrant, D; Marsh, G; Sorahan, T; and Weed DL. (2016). Glyphosate epidemiology expert panel review: a weight of evidence systematic review of the relationship between glyphosate exposure and non-Hodgkin's lymphoma or multiple myeloma. Crit Rev Toxicol. 2016 Sep;46(sup1):28-43.

Agricultural Health (cohort) Study. (2005). https://aghealth.nih.gov/.

Atkinson, C; Martin, T; Hudson, P; and Robb, D. (1993b). Glyphosate: 104-week dietary carcinogenicity study in mice. Inveresk Research International, Tranent, EH33 2NE, Scotland. IRI Project No. 438618. April 7, 1993. MRID 49631702. Unpublished.

Atkinson, C; Strutt, A; Henderson, W; et al. (1993a). 104-Week Chronic Feeding/ Oncogenicity study in rats with 52-week interim kill. MRID No. 49631701. Unpublished

Benedetti, D; Nunes, E; Sarmento, M; Porto, C; Iochims dos Santos, CE; Dias, JF; de Silva, J. (2013). Genetic damage in soybean workers exposed to pesticides: evaluation with the comet and buccal micronucleus cytome asays. Mutation Research. 752: 28-33. http://dx.doi.org/10.1016/j.mrgentox.2013.01.001.

Benjamini, Y and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B 57: 289-300.

Blair, A and Zahm, SH. (1993). Cancer among migrant and seasonal farmworkers: an epidemiologic review and research agenda. Am J Ind Med. 1993 Dec;24(6):753-66.

Bolognesi, C; Carrasquilla, G; Volpi, S; Solomon, KR; and Marshall, EJP. (2009). Biomonitoring of genotoxic risk in agricultural workers from five Colombian regions: Association to occupational exposure to glyphosate. Journal of Toxicology and Environmental Health - Part A: Current Issues 72, 986-997.

Brammer. (2001). Glyphosate Acid: Two Year Dietary Toxicity and Oncogenicity Study in Wistar Rats. Central Toxicology Laboratory, Alderley Park Macclesfield, Cheshire, UK: Syngenta. MRID 49704601. Unpublished.

Breslow, NE and Day, NE. (1980). Statistical methods in cancer research. Volume I - The analysis of case-control studies. IARC Sci Publ. (32):5-338.

Breslow, NE and Day, NE. (1987). Statistical methods in cancer research. Volume II--The design and analysis of cohort studies. IARC Sci Publ. (82):1-406.

Brown, LM; Burmeister, LF; Everett, GD; and Blair, A. (1993). Pesticide Exposures and Multiple Myeloma in Iowa Men. Cancer Causes Control 4, 153-156.

Chang, ET and Delzell, E. (2016). Systematic review and meta-analysis of glyphosate exposure and risk of lymphohematopoietic cancers. Journal of environmental science and health Part B, Pesticides, food contaminants, and agricultural wastes 51, 402-434.

Cocco, P. Satta, G. Dubois, S. Pili, C. Pilleri, M. Zucca, M. Mannetje, AM, Becker, N. Benavente, Y. de Sanjose, S. et al. (2013). Lymphoma risk and occupational exposure to pesticides: results of the Epilymph study. Occupational and environmental medicine 70, 91-98.

Cox, C and Surgan, M. (2006). Unidentified inert ingredients in pesticides: implications for human and environmental health. Environ Health Perspect. 2006 Dec;114(12):1803-6.

De Roos, AJ. Zahm, SH. Cantor, KP. Weisenburger, DD. Holmes, FF. Burmeister, LF. and Blair, A. (2003). Integrative assessment of multiple pesticides as risk factors for non-Hodgkin's lymphoma among men. Occupational and environmental medicine 60. 1-9.

De Roos, AJ; et al. (2005). Cancer incidence among glyphosate-exposed pesticide applicators in the Agricultural Health Study. Environ Health Perspect 113(1): 49-54.

Enemoto, K. (1997). HR-001: 24-Month Oral Chronic Toxicity and Oncogenicity Study in Rats, Vol. 1. The Institute of Environmental Toxicology, Kodaira-shi, Tokyo, Japan, Arysta Life Sciences, Study No.:IET 94-0150. MRID 50017104, 50017105, 5001703. Unpublished

EPA. (2010). EPA OPP Draft Framework for Incorporating Human Epidemiologic & Incident Data in Health Risk Assessment.

EPA. (2016a). Glyphosate Issue Paper: Evaluation of Carcinogenic Potential. EPA's Office of Pesticide Programs. September 12, 2016

EPA. (2016b). September 9, 2016 memorandum: Updated Statistics Performed on Animal Carcinogenic Study Data for Glyphosate. EPA-HQ-OPP-2016-0385-0095.

Eriksson, M; Hardell, L; Carlberg, M; and Akerman, M. (2008). Pesticide exposure as risk factor for non-Hodgkin lymphoma including histopathological subgroup analysis. International journal of cancer 123, 1657-1663.

Farmer, JH; Kodell, RL; and Greenman, DL. (1979). Dose and time response models for the incidence of bladder neoplasms in mice fed 2-acetylaminofluorene continuously. Journal of Environmental Pathology and Toxicology 3, 55-68.

Farrar DB and Crump KS (1988). Exact statistical tests for any carcinogenic effect in animal bioassays. Fundamental and Applied Toxicology 11: 652—663.

Ford, B. Bateman, LA. Gutierrez-Palominos, L. Park, R. Nomura, DK. (2017). Mapping Proteome-wide Targets of Glyphosate in Mice. Cell Chem Biol. 2017 Jan 19. pii: S2451-9456(16)30474-3. doi: 10.1016/j.chembiol.2016.12.013. [Epub ahead of print].

Gaines, TA; Zhang, WL; Wang, DF; Bukun, B; Chisholm, ST; Shaner, DL; Nissen, SJ; Patzoldt, WL; Tranel, PJ; Culpepper, AS; et al. (2010). Gene amplification confers glyphosate resistance in Amaranthus palmeri. P Natl Acad Sci USA 107, 1029-1034.

Gart, JJ; Krewski, D; Lee, PN; Tarone, RE; and Wahrendorf, J. (1986). Statistical methods in cancer research. Volume III--The design and analysis of long-term animal experiments. IARC Sci Publ. 1986;(79):1-219.

Greim, H; et al. (2015). Evaluation of carcinogenic potential of the herbicide glyphosate, drawing on tumor incidence data from fourteen chronic/carcinogenicity rodent studies." Crit Rev Toxicol 45(3): 185-208.

Hardell, L. Eriksson, M. and Nordstrom, M. (2002). Exposure to pesticides as risk factor for non-Hodgkin's lymphoma and hairy cell leukemia: pooled analysis of two Swedish case-control studies. Leuk Lymphoma. 2002 May;43(5):1043-1049.

Haseman, JK. (1983). Issues: a reexamination of false-positive rates for carcinogenesis studies. Fundam Appl Toxicol 3:334–339.

Haseman, JK. (1995). Data analysis: Statistical analysis and use of historical control data. Regul Toxicol Pharmacol 21:52–59.

Hoar, SK; Blair, A; Holmes, FF; Boysen, CD; Robel, RJ; Hoover, R; Fraumeni, JF Jr. (1986). Agricultural herbicide use and risk of lymphoma and soft-tissue sarcoma. JAMA. 256(9):1141-1147.

Hohenadel, K. Harris, SA. McLaughlin, JR. Spinelli, JJ. Pahwa, P. Dosman, JA. Demers, PA. and Blair, A. (2011). Exposure to multiple pesticides and risk of non-Hodgkin lymphoma in men from six Canadian provinces. International journal of environmental research and public health 8, 2320-2330.

Kachuri, L; Demers, PA; Blair, A; Spinelli, JJ; Pahwa, M; McLaughlin, JR; Pahwa, P; Dosman, JA; and Harris, SA. (2013). Multiple pesticide exposures and the risk of multiple myeloma in Canadian men. International journal of cancer 133, 1846-1858.

King-Herbert, A and Thayer, K. (2006). NTP workshop: animal models for the NTP rodent cancer bioassay: stocks and strains--should we switch? Toxicol Pathol. 34(6):802-805.

Knezevich, AL and Hogan, GK. (1983). A chronic feeding study of glyphosate in mice. Unpublished report prepared by Bio/Dynamic Inc., dated July 21, 1983. Report No. 77-2011. EPA Accession No. 251007 – 251009, and 251014. EPA Accession no. 251007-09, 251014. Unpublished.

Koureas, M; Tsezou, A; Tsakalof, A; Orfanidou, T; and Hadjichristodoulou, C. (2014). Increased levels of oxidative DNA damage in pesticide sprayers in Thessaly Region (Greece). Implications of pesticide exposure. The Science of the total environment 496, 358-364.

Koutros, S; Beane Freeman, LE; Lubin, JH; Heltshe, SL; Andreotti, G; Barry, K.H; DellaValle, CT; Hoppin, JA; Sandler, DP; Lynch, CF; et al. (2013). Risk of total and aggressive prostate cancer and pesticide use in the Agricultural Health Study. Am J Epidemiol 177, 59-74.

Kumar, DPS. (2001). Carcinogenicity Study with Glyphosate Technical in Swiss Albino Mice, Toxicology Department Rallis Research Centre, Rallis India Limited. Study No. TOXI: 1559.CARCI-M. MRID 49987403. Unpublished.

Lankas, GP. (1981). A Lifetime Study of Glyphosate in Rats. Report No. 77-2062 prepared by Bio Dynamics, Inc. EPA Accession. No. 247617 – 247621. December 23, 1981. MRID 00093879. Unpublished.

Lee, WJ; Colt, JS; Heineman, EF; McComb, R; Weisenburger, DD; Lijinsky, W; and Ward, MH. (2005). Agricultural pesticide use and risk of glioma in Nebraska, United States. Occup Environ Med. 2005 Nov; 62(11):786-92.

Marc, J; Mulner-Lorillon, O; Boulben, S; Hureau, D; Durand, G; and Belle, R. (2002). Pesticide Roundup® provokes cell division dysfunction at the level of CDK1/cyclin B activation. Chem Res Toxicol 15, 326-331.

Marc, J; Belle, R; Morales, J; Cormier, P; and Mulner-Lorillon, O. (2004). Formulated glyphosate activates the DNA-response checkpoint of the cell cycle leading to the prevention of G2/M transition. Toxicol Sci 82, 436-442.

McDuffie, HH; Pahwa, P; McLaughlin, JR; Spinelli, JJ; Fincham, S; Dosman, JA; Robson, D; Skinnider, LF; and Choi, NW. (2001). Non-Hodgkin's lymphoma and specific pesticide exposures in men: Cross-Canada study of pesticides and health. Cancer Epidemiol Biomarkers Prev 10, 1155-1163.

Mesnage, R; Defarge, N; Spiroux de Vendômois, J; and Séralini, GE. (2014). Major pesticides are more toxic to human cells than their declared active principles. Biomed Res Int. 2014;2014:179691. doi: 10.1155/2014/179691.

Orsi, L. Delabre, L. Monnereau, A. Delval, P. Berthou, C. Fenaux, P. Marit, G. Soubeyran, P. Huguet, F. Milpied, N. et al. (2009). Occupational exposure to pesticides and lymphoid neoplasms among men: results of a French case-control study. Occupational and environmental medicine 66, 291-298.

Pahwa, P; Karunanayake, CP; Dosman, JA; Spinelli, JJ; McDuffie, HH; and McLaughlin, JR. (2012). Multiple myeloma and exposure to pesticides: a Canadian case-control study. Journal of agromedicine 17, 40-50.

Pavkov, KLl. and Turnier, JC. (1987). Two-Year Chronic Toxicity and Oncogenecity Dietary Study with SC-0224 in Mice. Stauffer Chemical Company. MRID 40214006 and 41209907. Unpublished.

Pavkov, KLl. and Wyand, S. (1987). Two-Year Chronic Toxicity and Oncogenecity Dietary Study with SC-0224 in Rats. Stauffer Chemical Company. MRID 40214007, 41209905, and 41209907. Unpublished.

Portier, CJ. Armstrong. BK. Baguley, BC. Baur, X. Belyaev, I. Bellé, R. . . . Zhou, SF. (2016). Differences in the carcinogenic evaluation of glyphosate between the International Agency for Research on Cancer (IARC) and the European Food Safety Authority (EFSA). J Epidemiol Community Health. 2016 Aug; 70(8): 741–745. Published online 2016 Mar 3. doi: 10.1136/jech-2015-207005.

Reyna, MS. and Gordon, DE. (1973). 18-Month Carcinogenic Study with CP67573 in Swiss White Mice: IBT No. B569. (Unpublished study, including sponsor's validation report dated Feb 1, 1978, received Jun 21, 1978 under 524-308; prepared by Industrial Bio-Test Laboratories, Inc., submitted by Monsanto Co., Washington, D.C.; CDL:234136-G). MRID 00061113. Unpublished.

Rothman KJ. (1981). Induction and latent periods. Am J Epidemiol. 1981 Aug;114(2):253-259.

Schinasi, L and Leon, ME. (2014). Non-Hodgkin lymphoma and occupational exposure to agricultural pesticide chemical groups and active ingredients: a systematic review and meta-analysis. Int J Environ Res Public Health. 2014 Apr 23;11(4):4449-527. doi: 10.3390/ijerph110404449.

Schulz, KF and Grimes, DA. (2002). Case-control studies: research in reverse. Lancet. 2002 Feb 2;359(9304):431-4.

Séralini, G-E; Clair, E; Mesnage, R; Gress, S; Defarge, N; Malatesta, M; Hennequin, D; and Spiroux de Vendômois, J. (2014). Republished study: long-term toxicity of a Roundup® herbicide and a Roundup®-tolerant genetically modified maize. Environmental Sciences Europe; Bridging Science and Regulation at the Regional and European Level. 26:14 DOI: 10.1186/s12302-014-0014-5.

Sorahan, T. (2015). Multiple myeloma and glyphosate use: a re-analysis of US Agricultural Health Study (AHS) data. Int J Environ Res Public Health. 2015 Jan 28;12(2):1548-59. doi: 10.3390/ijerph120201548.

Stout, LD and Ruecker, PA. (1990). Chronic Study of Glyphosate Administered in Feed to Albino Rats. MRID No. 41643801; Historical Controls. MRID 41728700. Unpublished.

Sugimoto, K. (1997), HR-001: 18-Month Oral Oncogenicity Study in Mice, Vol. 1 and 2. The Institute of Environmental Toxicology, 2-772, Suzuki-cho, Kodaira-shi, Tokyo, 187, Japan, Study No.:IET 94-0151. MRID 50017108, 50017109. Unpublished.

Suresh, TP. (1996). Combined Chronic Toxicity and Carcinogenicity Study with Glyphosate Technical in Wistar Rats. Toxicology Department Rallis Research Centre, Rallis India Limited, TOXI-1559, 002/1-GPT-CARCI-M. MRID 49987401. Unpublished.

Tarone RE. (2016). On the International Agency for Research on Cancer classification of glyphosate as a probable human carcinogen. European Journal of Cancer Prevention; doi: 10.1097/CEJ.0000000000000289, Published ahead of print.

Viele, K; Verry, S; Neuenschwander, B; Amzal, B; Chen, F; Enas, N; Hobbs, B; Ibrahim, JG; Kinnersley, N; Lindborg, S; Micallef, S; Roychoudhury, S; and Thompson, L. (2014). Use of Historical Control Data for Assessing Treatment Effects in Clinical Trials. Pharm Stat. 13(1): 41–54. doi:10.1002/pst.1589.

Weisenburger, DD. (1992). Pathological classification of non-Hodgkin's lymphoma for epidemiologic studies. Cancer Res. 52:5456s–5464s.

Westfall P, Young S. (1989). P-value adjustments for multiple tests in multivariate binomial models. Journal of the American Statistical Association 84, 780-786.

Widholm, JM; Chinnala, AR; Ryu, JH; Song, HS; Eggett, T; and Brotherton, JE. (2001). Glyphosate selection of gene amplification in suspension cultures of 3 plant species. Physiol Plant 112, 540-545.

Williams, GM; Berry, C; Burns, M; de Camargo, JL; and Greim, H. (2016). Glyphosate rodent carcinogenicity bioassay expert panel review. Crit Rev Toxicol. 2016 Sep;46(sup1):44-55.

Wood, E; Dunster, J; Watson, P; and Brooks, P. (2009a). Glyphosate Technical: Dietary Combined Chronic Toxicity/Carcinogenicity Study in the Rat. Harlan Laboratories Limited, Shardlow Business Park, Shardlow, Derbyshire DE72 2GD, UK. Study No. 2060-012. April, 23, 2009. MRID 49957404. Unpublished.

Wood, E; Dunster, J; Watson, P; and Brooks, P. (2009b). Glyphosate Technical: Dietary Carcinogenicity Study in the Mouse. Harlan Laboratories Limited, Shardlow Business Park, Shardlow, Derbyshire DE72 2GD, UK. Study No. 2060-011. April, 22, 2009. MRID 49957402. Unpublished.

Zahm, SH; Weisenburger, DD; Babbitt, PA; Saal, RC; Vaught, JB; Cantor, KP; Blair. A. (1990). A case-control study of non-Hodgkin's lymphoma and the herbicide 2,4-dichlorophenoxyacetic acid (2,4-D) in Eastern Nebraska. Epidemiology. 1(5):349-356.

Zhang, C; Hu, R; Huang, J; Huang, X; Shi, G; Li, Y; Yin, Y; and Chen, Z. (2016). Health effect of agricultural pesticide use in China: implications for the development of GM crops. Sci. Rep. 6, 34918; doi: 10.1038/srep34918.

APPENDIX 1 – WRITTEN SUBMISSIONS TO DOCKET NO. EPA-HQ-OPP-2016-0385

Commenter*
Anonymous public comments
Comment submitted by Dag Falck, Organic Program Manager, Nature's Path Foods Inc.
Comment submitted by A. DeLuca
Comment submitted by A. Lewis
Comment submitted by A. Schneiderman
Comment submitted by A. Sorrells-Washington
Comment submitted by Aaron Hobbs, President, RISE, Responsible Industry for a Sound Environment
Comment submitted by Amechi Chukwudebe, PhD, BASF
Comment submitted by Amelia Jackson-Gheissari PhD, International Regulatory Affairs Manager and Donna Farmer, PhD, Senior Toxicologist, Monsanto Company
Comment submitted by Amy (no surname provided)
Comment submitted by Andrew Behar, Chief Executive Officer, As You Sow, et al.
Comment submitted by Anthony Samsel, Research Scientist, Consultant, Samsel Environmental and Public Health Services (SEAPHS)
Comment submitted by B. Talen
Comment submitted by B. Tarone
Comment submitted by Bill Freese, Science Policy Analyst, Center for Food Safety (CFS)
Comment submitted by C. A. Harris, PhD, FRSC et al.
Comment submitted by C. J. Portier, PhD
Comment submitted by C. Laieski
Comment submitted by Center for Regulatory Effectiveness (CRE)
Comment submitted by Christine T. (no surname provided)
Comment submitted by Christopher P. Wild, PhD, Director, International Agency for Research on Cancer (IARC)
Comment submitted by D. Brusick et al.
Comment submitted by D. Davis
Comment submitted by D. Norris
Comment submitted by D. Pompeo
Comment submitted by D. Schubert
Comment submitted by D. Sutherland
Comment submitted by Dale Moore, Executive Director, Public Policy, American Farm Bureau Federation (AFBF)
Comment submitted by Daniele Court Marques, Pesticides Unit, European Food Safety Authority (EFSA)
Comment submitted by Danielle (no surname provided)
Comment submitted by David Spak, Stewardship Manager, Bayer Vegetation Management
Comment submitted by Deborah Larson Hommer, President, Virginians for Medical Freedom (VMF)
Comment submitted by Dennis D. Weisenburger, MD, Professor, Chair, Department of Pathology, City of Hope

Commenter*
Comment submitted by Dev Gowda, J.D., Toxics Advocate, U.S. Public Interest Research Group (PIRG)
Comment submitted by Donna Farmer, PhD. Senior Toxicologist, Monsanto Company
Comment submitted by Dow AgroSciences
Comment submitted by E. Crouch
Comment submitted by E. Springwind
Comment submitted by E. Stockman
Comment submitted by E. Wilson
Comment submitted by Emily Marquez, Ph.D., Staff Scientist, Pesticide Action Network North America (PANNA)
Comment submitted by G. Stromberg
Comment submitted by Gordon Stoner, President, National Association of Wheat Growers (NAWG)
Comment submitted by Gretchen DuBeau, Esq., Executive and Legal Director, Alliance for Natural Health USA
Comment submitted by H. Rowland
Comment submitted by I. Panton
Comment submitted by Intertek Scientific & Regulatory Consultancy
Comment submitted by J. Hoy
Comment submitted by J. Littrell
Comment submitted by J. Manning
Comment submitted by J. Moore
Comment submitted by J. Young
Comment submitted by James S. Bus PhD, DABT, ATS, Exponent, Inc. on behalf of CropLife America
Comment submitted by Janet E. Collins, Ph.D., R.D., Senior Vice President, Science and Regulatory Affairs, CropLife America (CLA)
Comment submitted by Jennifer Sass, PhD, Senior Scientist, Natural Resources Defense Council (NRDC)
Comment submitted by John Weinand, President, North Dakota Grain Growers Association (NDGGA)
Comment submitted by Joseph K. Haseman, J. K. Haseman Consulting
Comment submitted by K. Lundsford
Comment submitted by K. Taylor
Comment submitted by Kevin Bradley, President, Weed Science Society of America (WSSA)
Comment submitted by L. Garvey
Comment submitted by L. Staman
Comment submitted by Lars Niemann, Toxicology of Active Substances and their Metabolites Unit, Department Safety of Pesticides, German Federal Institute for Risk Assessment (BfR), Berlin
Comment submitted by Luther Markwart, Executive Vice President, American Sugarbeet Growers Association and Co-Chairman, Sugar Industry Biotechnology Council
Comment submitted by M. Bosland

Commenter*
Comment submitted by M. McLean
Comment submitted by M. Moore
Comment submitted by M. Pybus
Comment submitted by M. Wilkus
Comment submitted by Montague Dixon, Senior Regulatory Manager, Syngenta Crop Protection, LLC
Comment submitted by Ms. Delgado
Comment submitted by N. Paffrath
Comment submitted by Nathan Donley, Ph.D., Senior Scientist, Environmental Health Program, Center for Biological Diversity
Comment submitted by Nichelle Harriott, Science and Regulatory Director, Beyond Pesticides
Comment submitted by Nufarm Americas Inc
Comment submitted by P. A. Fenner-Crisp
Comment submitted by P. Whitman
Comment submitted by Pamela Koch, EdD, RD, Executive Director Laurie M. Tisch Center for Food, Education & Policy, Teachers College, Columbia University
Comment submitted by Peter F. Infante, Consultant, Peter F. Infante Consulting, L.L.C.
Comment submitted by Philip W. Miller, Ph.D., Vice President, Global Regulatory and Government Affairs, Monsanto Company
Comment submitted by R. Andrews
Comment submitted by R. Briggs
Comment submitted by R. E. Tarone
Comment submitted by R. Mason
Comment submitted by R. Parsons
Comment submitted by R. Tarone
Comment submitted by Rebecca St James (no surname provided)
Comment submitted by Reece Langley, Vice President, Washington Operations, National Cotton Council (NCC)
Comment submitted by Richard D. Gupton, Senior Vice President, Public Policy & Counsel, Agricultural Retailers Association (ARA)
Comment submitted by Richard Wilkins, President, American Soybean Association (ASA)
Comment submitted by Robert P. DeMott, Principal Toxicologist, Ramboll Environ on behalf of The Scotts Company LLC
Comment submitted by S. Seneff
Comment submitted by S. Barr
Comment submitted by S. Gardon
Comment submitted by S. Stair
Comment submitted by S. Vose
Comment submitted by S. Young
Comment submitted by Scott Slaughter, The Center for Regulatory Effectiveness (CRE)
Comment submitted by Steve Levine, Ph.D., CropLife America (CLA)
Comment submitted by T. Tokuda

Commenter*
Comment submitted by Tony Tweedale, R.I.S.K. (Rebutting Industry Science with Knowledge) Consultancy
Comment submitted by W. Beck
Comment submitted by W. Fawell
Comment submitted by Wenonah Hauter, Executive Director, Food & Water Watch
Comment submitted by Zen Honeycutt, Executive Director, Moms Across America
Mass Comment Campaign submitted by Alexis Baden-Mayer, Organic Consumers Association
Mass Comment Campaign submitted by Jennifer Listello, Existing Chemistry Global Coordinator, Monsanto
Mass Comment Campaign submitted by Tiffany Finck-Haynes, Friends of the Earth
Mass Comment Campaign submitted by Alliance for Natural Health USA
Mass Comment Campaign submitted by Food and Water Watch
Mass Comment Campaign submitted by Tiffany Finck-Haynes, Friends of the Earth

*Note: some commenters provided multiple submissions to the docket.