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

OFFICE OF CHEMICAL SAFETY
AND POLLUTION PREVENTION

December 9, 2010

MEMORANDUM

SUBJECT: Transmittal of Meeting Minutes of the FIFRA Scientific Advisory Panel Meeting held September 14 - 17, 2010 on the Re-evaluation of the Human Health Effects of Atrazine: Review of Non-cancer Effects and Drinking Water Monitoring Frequency

TO: Steven Bradbury, Ph.D.
Director
Office of Pesticide Programs

FROM: Joseph E. Bailey, Designated Federal Official
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

THRU: Laura Bailey, Executive Secretary
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

Frank Sanders, Director
Office of Science Coordination and Policy

Attached, please find the meeting minutes of the FIFRA Scientific Advisory Panel open meeting held in Arlington, VA on September 14 - 17, 2010. This report addresses a set of scientific issues associated with the Re-evaluation of the Human Health Effects of Atrazine: Review of Non-cancer Effects and Drinking Water Monitoring Frequency.

Enclosure

cc:

Stephen Owens
James J. Jones
Vicki Dellarco
William Jordan
Margie Fehrenbach
Keith Matthews
Donald Brady
William Diamond
Jack Housenger
Tina Levine
Joan Harrigan-Farrelly
Lois Rossi
Robert McNally
Richard Keigwin
Anna Lowit
Elizabeth Mendez
John Liccione
Robert Luebke
Carol Christensen
Chester Rodriguez
Ralph Cooper
Mary Frankenberry
Nelson Thurman
Enesta Jones
Douglas Parsons
Vanessa Vu
OPP Docket

FIFRA Scientific Advisory Panel Members

Steven Heeringa, Ph.D. (FIFRA SAP Chair)
John Bucher, Ph.D., DABT
Janice Chambers, Ph.D., DABT, Fellow ATS
Gerald LeBlanc, Ph.D.
Kenneth Portier, Ph.D.
Daniel Schlenk, Ph.D.

FQPA Science Review Board Members

Susan F. Akana, Ph.D.
John Bailar, III, M.D., Ph.D.
Richard H. Coupe, Ph.D.
Kenneth Barry Delclos, Ph.D.
Penelope Fenner-Crisp, Ph.D., DABT
Ellen Gold, Ph.D.
Richard Greenwood, Ph.D.
Shelley Harris, Ph.D.
Nelson Horseman, Ph.D.
Kannan Krishnan, Ph.D.
Herbert K. H. Lee, Ph.D.
Sandra Legan, Ph.D.
James McManaman, Ph.D.
Bette Meek, Ph.D.
Moiz Mumtaz, Ph.D.
Katherine Roby, Ph.D.
Wesley Stone, M.S.

SAP Minutes No. 2010-07

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

Re-evaluation of the Human Health Effects of Atrazine: Review of Non-Cancer Effects and Drinking Water Monitoring Frequency

September 14 – 17, 2010
FIFRA Scientific Advisory Panel Meeting
Held at the
Environmental Protection Agency Conference Center
Arlington, VA

NOTICE

These meeting minutes have been written as part of the activities of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP). The meeting minutes represent the views and recommendations of the FIFRA SAP, not the United States Environmental Protection Agency (Agency). The content of the meeting minutes does not represent information approved or disseminated by the Agency. The meeting minutes have not been reviewed for approval by the Agency and, hence, the contents of these meeting minutes do not necessarily represent the views and policies of the Agency, nor of other agencies in the Executive Branch of the Federal Government, nor does mention of trade names or commercial products constitute a recommendation for use.

The FIFRA 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, Office of Pesticide Programs (OPP), and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. FQPA Science Review Board members serve the FIFRA SAP on an ad hoc basis to assist in reviews conducted by the FIFRA SAP. Further information about FIFRA SAP reports and activities can be obtained from its website at http://www.epa.gov/scipoly/sap/ or the OPP Docket at (703) 305-5805. Interested persons are invited to contact Joseph E. Bailey, SAP Designated Federal Official, via e-mail at bailey.joseph@epa.gov.

In preparing these meeting minutes, the Panel carefully considered all information provided and presented by EPA, as well as information presented by public commenters.

TABLE OF CONTENTS

SAP Minutes No. 2010-07

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

Re-evaluation of the Human Health Effects of Atrazine: Review of Non-Cancer Effects and Drinking Water Monitoring Frequency

September 14 – 17, 2010
FIFRA Scientific Advisory Panel Meeting
Held at the
Environmental Protection Agency Conference Center
Arlington, VA

Steven G. Heeringa, Ph.D.
FIFRA SAP Chair
FIFRA Scientific Advisory Panel
Date: December 9, 2010

Joseph E. Bailey
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: December 9, 2010

Federal Insecticide Fungicide and Rodenticide Act
Scientific Advisory Panel Meeting
September 14 - 17, 2010

PARTICIPANTS

FIFRA SAP Chair

Steven G. Heeringa, Ph.D., Research Scientist & Director for Statistical Design, University of Michigan, Institute for Social Research, Ann Arbor, MI

Designated Federal Official

Joseph E. Bailey, FIFRA Scientific Advisory Panel, Office of Science Coordination and Policy, EPA

FIFRA Scientific Advisory Panel Members

John Bucher, Ph.D., DABT, Associate Director, National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC

Janice Chambers, Ph.D., DABT, Fellow ATS, Director, Center for Environmental Health Sciences, College of Veterinary Medicine, Mississippi State University, Mississippi State, MS

Gerald A. LeBlanc, Ph.D., Professor and Department Head, Department of Environmental & Molecular Toxicology, North Carolina State University, Raleigh, NC

Cary N. Pope, Ph.D., Professor, Head & Sitlington Chair of Toxicology, Department of Physiological Sciences, Oklahoma State University College of Veterinary Medicine, Stillwater, OK

Kenneth Portier, Ph.D., Program Director, Statistics, American Cancer Society, National Home Office, Atlanta, GA

Daniel Schlenk, Ph.D., Professor of Aquatic Ecotoxicology and Environmental Toxicology, Department of Environmental Sciences, University of California, Riverside, CA

FIFRA Science Advisory Board Members

Susan F. Akana, Ph.D., Associate Professional Researcher, Department of Physiology University of California San Francisco, San Francisco, CA

John C. Bailar, III, M.D., Ph.D., Scholar in Residence, The National Academies, Professor Emeritus, University of Chicago, Washington, DC

Richard H. Coupe, Ph.D., Supervisory Hydrologist/Associate Water Science Director, U.S. Geological Survey, Pearl, MS

Kenneth Barry Delclos, Ph.D., Pharmacologist, FDA, National Center for Toxicological Research, Jefferson, AR

Penelope A. Fenner-Crisp, Ph.D., DABT, Consultant, North Garden, VA

Ellen B. Gold, Ph.D., Professor and Chair, Department of Public Health Sciences, School of Medicine, University of California - Davis, Davis, CA

Richard Greenwood, Ph.D., Professor of Environmental Science, University of Portsmouth, Portsmouth, UNITED KINGDOM

Shelley A. Harris, Ph.D., Scientist, Prevention and Cancer Control, Cancer Care Ontario, Associate Professor, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, CANADA

Nelson D. Horseman, Ph.D., Professor, Department of Molecular and Cellular Physiology, Department of Medicine (Endocrinology & Metabolism), Program in Systems Biology and Physiology, University of Cincinnati, Cincinnati, OH,

Kannan Krishnan, Ph.D., Professor, Department of Occupational and Environmental Health, University of Montreal, Montreal, CANADA

Herbert K. H. Lee, Ph.D., Professor, Department of Applied Mathematics and Statistics, Associate Dean for Graduate Studies and Research, Baskin School of Engineering, University of California - Santa Cruz, Santa Cruz, CA

Sandra J. Legan, Ph.D., Professor, Department of Physiology and Biophysics, University of Kentucky, Lexington, KY

James L. McManaman, Ph.D., Professor, Department of Obstetrics and Gynecology, Division Chief, Basic Reproductive Sciences, University of Colorado – Denver, Aurora, CO

M.E. Bette Meek, Ph.D., Associate Director, Chemical Risk Assessment, McLaughlin Centre for Population Health, University of Ottawa, Ottawa, Ontario, CANADA

Moiz M. Mumtaz, Ph.D., Science Advisor/Senior Toxicologist, Division of Toxicology and Environmental Medicine, Agency for Toxic Substances & Disease Registry, Centers for Disease Control & Prevention, Atlanta, GA

Katherine Roby, Ph.D., Director, Reproductive Endocrinology Laboratory, Center for Advanced Reproductive Medicine, University of Kansas Medical Center, Kansas City, KS

Wesley W. Stone, M.S., Hydrologist, U.S. Geological Survey, Department of Interior, Indianapolis, IN

INTRODUCTION

The Federal Insecticide, Fungicide and Rodenticide Act Scientific Advisory Panel (FIFRA SAP) has completed its review of the Reevaluation of the Human Health Effects of Atrazine: Review of Non-cancer Effects and Drinking Water Monitoring Frequency. Advance notice of the meeting was published in the *Federal Register* on June 11, 2010. The review was conducted in an open panel meeting held in Arlington, VA, on September 14 - 17, 2010. Dr. Steven Heeringa chaired the meeting. Joseph E. Bailey served as the Designated Federal Official.

EPA is undertaking a re-evaluation of the human health effects of atrazine. The re-evaluation has involved three SAP meetings in 2010. The first was held in February 2010 at which the Agency presented its preliminary reviews of several epidemiologic studies of the relation of atrazine to birth outcomes and described a plan to evaluate atrazine epidemiology data from the Agricultural Health Study. The second SAP meeting, held in April 2010, focused on 1) a preliminary review of experimental toxicology studies from laboratory mammals and *in vitro* studies and recent advancements in understanding atrazine's mode of action along with 2) statistical approaches for evaluating monitoring frequency in community water systems (CWS). The third SAP in September 2010 focused on the non-cancer effects of atrazine. The Agency included studies available up through July 15, 2010. At this meeting OPP integrated data from *in vitro* and *in vivo* experimental toxicology studies along with preliminary review of non-cancer epidemiologic studies in a draft weight of the evidence (WOE) analysis. The Agency will use feedback received from the SAP at the September 2010 meeting as it completes the non-cancer WOE analysis integrating experimental toxicology and epidemiology studies with statistical analysis for determining whether or not adjustments are necessary in the sampling frequency of CWS water monitoring.

Opening remarks at the meeting were provided by Steven Bradbury, Ph.D., Director, Office of Pesticide Programs and Tina Levine, Ph.D., Director, Health Effects Division, Office of Pesticide Programs. Agency presentations were given by Anna Lowit, Ph.D., Carol Christensen, Ph.D. and Chester Rodriguez, Ph.D., Health Effects Division; Nelson Thurman, M.S., and Mary Frankenberry, Environmental Fate and Effects Division, all of the Office of Pesticide Programs. In addition, a presentation was given by Suzanne Fenton, Ph.D., of the National Institute of Environmental Health Sciences, National Institutes of Health.

PUBLIC COMMENTS

Oral Statements were presented as follows:

Janis McFarland, Ph.D., Charles Breckenridge, Ph.D., Russell C. Hovey, Ph.D. (University of California - Davis), Robert Handa, Ph.D. (The University of Arizona), James Simpkins, Ph.D. (University of North Texas), David Kim, Ph.D., Tony M. Plant, Ph.D. (University of Pittsburgh) and Paul Hendley, Ph.D. all on behalf of Syngenta Crop Protection

Tyrone B. Hayes, Ph.D. (University of California - Berkeley) on behalf of himself and other scientists

Rod Snyder on behalf of the National Corn Growers Association

Jere White, James C. Lamb, Ph.D., Gary Marshall, John Hall and
David C. Bridges, Ph.D. (Abraham Baldwin Agricultural College) all on behalf of the Triazine Network

Scott Slaughter on behalf of the Center for Regulatory Effectiveness

Tyler Wegmeyer on behalf of the American Farm Bureau Federation

Rebeckah Freeman Adcock on behalf of Croplife America

Written Statements were provided by:

Jean Public
Dan Campbell on behalf of Syngenta Crop Protection
Scott Slaughter on behalf of the Center for Regulatory Effectiveness
Tyrone B. Hayes on behalf of himself and other scientists
Michael Leggett, Ph.D., on behalf of CropLife America

Summary of Panel Discussion and Recommendations

Charge Issue 1.1: Non-Cancer Epidemiology Reviews - EPA's critique of the atrazine epidemiology literature related to non-cancer outcomes is thoughtful and comprehensive; however, the Panel did not agree with EPA's conclusion that the database for non-cancer outcomes is strong. Of 29 candidate papers, 19 met the Agency's criteria for evaluation. They were spread over many different health outcomes and had many important limitations as noted by EPA and the Panel's detailed analysis. The Panel had the following comments regarding the methodology used to evaluate the papers and recommendations: 1) the criteria for including studies should be described in greater detail; 2) papers that were excluded from full evaluation should be discussed in an appendix with the rationale for their exclusion; 3) two of the 29 candidate papers were excluded primarily because they did not independently measure atrazine or triazines, but this was not well justified by EPA; 4) the 46 exposure assessment studies identified in the literature search should be reviewed and summarized in any future cancer risk evaluation, and it was suggested that the review and summary be provided to the SAP for consideration; 5) a system of scoring study quality should be developed and each study should be independently scored by at least two reviewers; 6) ecologic studies should rarely be used for evaluating weight of evidence and are useful only for purposes of hypothesis generation; and 7) additional studies other than those reviewed have assessed some of the outcomes better and have stored biologic samples and thus have the potential to examine the relation of atrazine to non-cancer outcomes more effectively, and the EPA should pursue measurement of atrazine and analyses of the existing outcomes in these studies. In general, the Panel believed the 19 non-cancer outcome studies reviewed were suggestive of possible adverse health outcomes, but had limitations that need to be addressed more fully.

Charge Issue 1.2a & b: Non-Cancer Epidemiology Database & Risk Assessment - Overall, the Panel agreed with EPA's conclusions that the current non-cancer epidemiologic database provides some useful information for hazard identification but is too limited to provide sufficient information for credible dose-response assessment or risk characterization. To integrate epidemiology and experimental toxicology information, the following are needed: 1) a strong experimental toxicology database; 2) understanding of toxicological mode(s) of action; and 3) a strong epidemiology database. For atrazine, a robust experimental toxicology database exists and understanding its modes of action is increasing. However, with regard to human epidemiologic studies, studies are needed that are specifically designed to evaluate relevant outcomes and that are based on defined mode(s) of action. The 19 epidemiology studies evaluated presented very little information collected in humans that support the outcomes observed in animals; therefore, the Panel concluded none of the non-cancer epidemiology studies were useful other than for problem formulation. The Panel agreed with EPA's conclusion that the experimental toxicology data and not the epidemiologic data are best suited for dose-response and risk assessment at the present time, but that the epidemiologic and toxicologic databases, in combination, can be used to inform hazard identification and characterization. Contrary to the Agency's opinion stated in the Issue Paper, the Panel

believed that epidemiologic findings could be used for endpoint selection if the studies are well-designed and experimental data are not available.

Charge Issue 2.1: Mammary Gland Development Studies (Rayner *et al*., 2004, 2005 and Coder, 2010) - The Panel pointed out important experimental design and methodology differences between the Rayner and Coder studies related to the following aspects: 1) composition of the diet consumed by the dams; 2) light-dark photoperiod employed; 3) inclusion/absence of pair-fed controls; 4) methods used for mammary gland development analysis; and 5) differences in maternal stress levels. The Panel concluded that these methodological and procedural differences confounded the understanding of the toxicological significance of exposure to atrazine on certain aspects of female reproductive system development. It was the Panel's opinion that unless/until these differences are resolved, an understanding of the effects of atrazine on mammary gland development will be hampered. Consequently, the Panel believed that use of the existing data on rat mammary gland development to assess the potential human risk of atrazine is not warranted at the present time.

Charge Issue 2.2 & 2.3: Enoch (2007) Mammary Gland Study and Use of Rayner, Coder and Enoch Studies in Hazard Assessment - The Panel agreed with EPA's conclusion that the Enoch study is not suitable for use in quantitative risk assessment. However, the study does raise some interesting points and dilemmas. Certainly, the very low doses at which effects on mammary gland development were reported could be of importance in assessing the risks of atrazine exposure. However, since these effects occurred at doses so much lower than observed in other studies, replication of the study is needed. The observations of this study have not been corroborated using other mixtures of parent atrazine and metabolites reflective of temporal and spatial occurrence or for atrazine parent at equivalent doses. Atrazine itself will, once absorbed, yield a mixture of free and conjugated metabolites (but little or no hydroxyatrazine) and the proportion of metabolites would likely change temporally. Any replication would need to evaluate the same endpoints measured in this study and, perhaps, other physiological endpoints. The Panel concurred with the Agency's conclusion that the few available mammary gland studies do not constitute a sufficient body of evidence to be used in a quantitative risk assessment (i.e., for establishing a Point of Departure (PoD)). However, the Panel was clear in stating that this conclusion should not be construed to mean that the mammary gland is not a legitimate target for further analysis and concluded that the suggestive evidence in these studies argues strongly for development of useful and standardized methods for assessing mammary gland functional outcomes.

Charge Issue 3.1a & b: Thede (1987) Pharmacokinetic (PK) Data and PK Pseudo-Steady State - The Panel believed it would be prudent for the Agency to work with the current PBPK model described by the registrant and make appropriate modifications. The Panel recommended that the Agency be more proactive in this area, and be more rigorous in addressing the uncertainties of extrapolation from rodent studies to low dose regimens in humans. However, in the absence of a functional PBPK model and appropriate data, the Panel strongly supported the work of the Agency in pursuing a dose-response analysis based on an internal dose metric, as an alternative to the administered dose metric used in

the 2003/2006 risk assessment. The Panel agreed with the Agency that, on the basis of the currently available data, plasma appears to be a reasonable biological compartment that is reflective of tissue dose, and that use of area under the plasma concentration time curve (AUC) provides an appropriate measure of internal exposure. It is the best way of approaching the problem of relating the exposure of rats to atrazine to the magnitude of resulting biological responses, and extrapolating to the prediction of human exposure levels. In the absence of good data on the plasma concentrations of parent compound and individual metabolites at various elapsed times after dosing, the use of total chlorotriazine based on total ^{14}C-compounds is a reasonable first step (particularly in the absence of information on the pharmacodynamic activity of the parent compound and individual metabolites). However, the Panel believed that the Thede (1987) data, had limitations, and should be interpreted with caution.

In the opinion of the Panel, the approach taken by the Agency is correct, but requires better data than those currently available. The pharmacokinetic study in progress as described in the registrant's presentation at the meeting and in Breckenridge, *et al*. (2010) will provide data of the required quality since the use of frequent sampling and mass spectrometric detection of the analytes is providing tight definitions of the PK profiles of the individual compounds (parent compound and analytes). This type of study lends itself to analysis of relationships between the AUCs of the different compounds and biological responses. This information combined with good estimates of plasma partition coefficients should allow the development of physiologically based pharmacokinetic models that can be used to extrapolate from rodents to other species, including humans. This would provide a sound basis for the setting of benchmark doses, and PoDs, providing that it is developed for the rat strain in which the PoD or dose-response is characterized.

Pseudo steady state implies that the absorption and distribution processes are matched by the elimination processes. The interpretation of the levels presented in Figure 5.5 of the Issue Paper is misleading since all plasma samples were taken at 24 h after the previous dose. The points defining the profiles in Figure 5.5 are the minima in the sawtooth profile, and not the average levels. Perhaps if the samples had been taken at 12 h after dosing, the profile would have been different. If the samples had been taken at random times after dosing, there would not have been such a smooth plateau. The real profile will not have such small amplitude fluctuations as illustrated in Figure 5.4 of the Issue Paper (see Figure 6 in Breckenridge, *et al*. (2010)). This means that the AUC is underestimated. The bias will be approximately constant across the plateau area, and so it will not change the goodness of fit of AUC versus administered dose, and could give a false sense of security.

Charge Issue 3.1c: PK Pseudo-steady State and Attenuation of LH Surge - The Panel was asked to comment on the Agency's analysis and preliminary conclusions related to Figure 5.8 in the Issue Paper which shows a plot of LH attenuation data versus administered atrazine across a range of exposure durations. The Panel consensus was that Figure 5.8 is an imaginative distillation of data from two different rat strains across 15 years of data collection from four laboratories. The Figure serves as a visual meta-

analysis and assists in the understanding of the atrazine dose-related attenuation of the LH surge. It is remarkable that the overall response is similar given the relatively great diversity in the study paradigms: different rat strains and different dosing regimens. The model presented in Figure 5.8 represents a good, initial approach to data integration and has great utility for exploring concepts and formulating approaches to test hypotheses further. The PK data indicated that a pseudo-steady state was achieved over the four days of dosing in the Thede experiments. However, it is important that interpretation of these data not use the idea that a pseudo-steady state is necessary. The data do not support this interpretation because no other dosing pattern was used. Overall, the Panel agreed that several aspects of Figure 5.8 need refinement in terms of the axis scaling, winnowing of LH data, inclusion of the higher doses to incorporate the full range of data and the separate evaluation of data from two rat strains. All these aspects can be corrected and may yield a more accurate and sensitive model.

Charge Issue 3.2a: Benchmark Dose Analysis - The Panel agreed with the Agency's conclusion that, based on the available data, a benchmark dose (BMD) modeled from data on suppression of the LH surge appears to be protective for other endpoints, since this phenomenon occurs at doses lower than for the wide range of effects identified in a rather extensive toxicological database. While attenuation of the LH surge over one estrous cycle is a sensitive biological response, a causal association between the modest, yet statistically significant, observed changes in LH surge peak and adverse fertility measures has not been made. The relatively small attenuation of the LH surge may well be a harbinger of a possible future adverse event, but at present this remains to be demonstrated and other possible endpoints should continue to be investigated.

The biological significance and degree of adversity of the effect used to derive the benchmark value remains to be addressed robustly and needs to be considered carefully in relation to other endpoints on continued exposure, based on the totality of data, including that from more traditional studies. It may be helpful, then, to develop a variety of potential points of departure for relevant endpoints with a view to characterizing their degree of protection (in relation to biological significance of the observed effect) and associated uncertainty. This would serve as a critical step for the interpretation of the derived BMD, not just to provide understanding of the biological significance of observed effects, but also as a tool to compare and contrast the uncertainties associated with various options. Comparative uncertainty analysis would include consideration of appropriate interspecies and intraspecies adjustments, in the context of their associated degree of uncertainty.

The existing data on the dose response of the attenuation of the LH surge associated with exposure to atrazine (in rats) are sufficient to permit robust analysis for the benchmark dose. These include significant data in several strains exposed to a range of doses delivered either by oral gavage or in the diet. The Agency modeled four relevant candidate data sets, representing the four studies summarized in Figure 5.8 of the Issue Paper and selected the “new” NHEERL 4-day data (Cooper 2010, unpublished) as the most appropriate for deriving a PoD. Selection of this study as the basis for BMD modeling is appropriate, not only because of the range of doses used, and the lower

variability of the data, but also owing to the inclusion of data on animals exhibiting a statistically defined LH surge.

C**harge Issue 3.2b: Benchmark Response for LH Attenuation -** A solid understanding of a benchmark response is needed for BMD analysis. In this case, attenuation of the LH surge is a sensitive measure of atrazine exposure. It directly relates to mode of action, being one of the key precursor events to the ultimate endpoint(s) of toxicity, and it exhibits dose response characteristics. Thus, it is a suitable benchmark response. Even so, inherent difficulties exist. These include the wide range in LH surge amplitude normally observed in both rodents and women, and the variability in the exact time of day the actual peak in LH surge occurs. There are ample data in the literature to demonstrate that complete (or nearly complete, i.e. ~ 80%+) inhibition of the LH surge, does directly correlate with adverse fertility outcomes, including altered cyclicity and inhibition of ovulation. Outside of these very dramatic changes in LH surge (at or near complete abrogation), data correlating more modest, yet statistically significant, changes in LH surge characteristics to key fertility events are lacking. These data are not only lacking from the series of studies presented in the Issue Paper, but importantly, there are not sufficient data in the broad body of literature (rodent or human) related to reproduction and fertility to directly link small or moderate changes in LH surge characteristics with fertility outcomes. This does not negate the possibility that they do exist.

Based on the Panel's current understanding of effects related to atrazine exposure together with its level of understanding of long-term effects of modest changes in LH surge and fertility, they concluded that the available data are insufficient to directly relate a specific percent change in either amplitude or AUC of the LH surge with fertility outcomes. At the present time, use of data on attenuation of the LH surge is a suitable benchmark response and directly relates to the present understanding of mode of action. However, additional endpoints, including but not limited to growth hormone activity, should not be overlooked. Furthermore, consideration of additional endpoints and/or additional modes of action may allow for integration of both cancer and non-cancer health effects of atrazine.

Charge Issue 4.1: EPA's Framework for Designing Monitoring Study - Regarding EPA's recommended framework for designing a drinking water monitoring study, the Panel noted that if the true purpose of a monitoring program is to generate valid “human exposure estimates,” then sampling the finished water of a community water system (CWS) is more appropriate than sampling its raw input water. EPA requested the Panel to specifically comment on four elements of its recommended framework for designing a monitoring study. The Panel agreed that targeting the most vulnerable areas is appropriate; however, caution was urged in how areas are defined as “vulnerable.” Frequency of sampling and selection of criteria for inclusion in the monitoring program should be carefully considered to avoid underestimating true atrazine concentrations. The Panel noted that monitoring programs need flexibility to accommodate year to year variability in agricultural practices. The Panel agreed with EPA that it is appropriate to sample intensively during periods expected to have high occurrence of exceedances. To help predict occurrence, some Panelists suggested use of an equation that considers

atrazine application timing and amounts along with hydrology of the site. Further, the Panel indicated that reservoirs warranted separate consideration when designing a monitoring program. Basing sampling frequency on the toxicological exposure duration of concern, as proposed by EPA, would be appropriate if there was agreement on what constitutes toxicological exposures of concern. Some Panel members expressed the opinion that there was not strong evidence of adverse health impacts from atrazine exposure at levels currently observed in surface waters. On the other hand, the fact that atrazine is so widely used and found in many environmental compartments raises the importance of further monitoring. The Panel agreed that it is important to continue to monitor atrazine to better understand exposure levels and to ensure that exposure levels continue to decline. Finally, the Panel agreed with EPA that autosamplers are appropriate to use to collect data for exposure periods of interest when quality assurance and quality control programs are in place to ensure reliable information.

Charge Issue 4.2a: Use of Chemograph Shapes to Match CWS with Intensively Monitored Datasets - Given the likelihood of missing short-duration concentration peaks and the variability of chemograph shapes over time at a site, using the chemograph shape as the mechanism to link CWS sites and the high sampling frequency sites, may be overly difficult and may provide a false sense of confidence in terms of assessing the sampling frequency adequacy for a site. Functional Data Analysis (Ramsey and Silverman, 2002; 2005) was suggested as one approach to grouping sites based on chemograph shapes; however, care must be taken in accounting for missing peak concentrations in this analysis. As an alternative approach, the Panel suggested matching high frequency sampling sites to community water supply sites based on water body and watershed characteristics.

Charge Issue 4.2b: Heidelberg and AEEMP Datasets for Analysis of Flowing Water and PRZM/EXAMS for Lakes and Reservoirs - When developing methodology, it is important to have true daily data, as both the Heidelberg and AEEMP datasets do. The Heidelberg dataset is of high quality and reflects frequent sampling, but it does not represent geographical diversity. Therefore, generalization from this dataset to the data from other watersheds would need to be done with care to avoid less than accurate predictions. The Perry Lake dataset would be more relevant to reservoirs. PRZM/EXAMS modeling is also useful, providing calibrated curves with simulated daily values; however, it should be noted that these models provide approximations of true values and that appropriate caution should be used in drawing conclusions from these data. PRZM/EXAMS is a very detailed model, not the easiest to use in practice, and needs to be tailored to each individual watershed. There was some sentiment among the Panel members that a more practical use of resources could be the collection of more varied, intensely-sampled datasets (particularly for reservoirs), rather than attempting to fit complex, labor-intensive, site-specific models to less frequently-sampled datasets. Data from intensely sampled sites would be expected to give much more accurate results than a linear or stair-step interpolation of less frequently sampled data (such as weekly data). The Panel noted that there is substantial literature on the use of computer models to augment physical data, and when properly calibrated (i.e., adjusted to match the available sparser real data), the approximations can be very good.

Charge Issue 4.3: Regression-based Modeling Combined with Random Function Modeling - Combining a deterministic model, like SEAWAVE-Q or PRZM/EXAMS, with a regression-based model, like WARP, appears to be the most promising approach to deal with sparse data in flowing water (the scenario for which these models were designed). This approach makes use of information about the pathways and aggregate data that are not incorporated by purely statistical methods such as kriging or neural networks. Combining a model such as WARP with a purely statistical method will give better interpolation than simple methods such as linear interpolation, but it would fail to make use of information about the expected shape of concentration curves that could be predicted by a deterministic model. However, there is a trade-off in the difficulty of implementing these methods, and it may be easier to implement a combination of WARP and a relatively simple statistical approach such as kriging, compared with combining WARP with a deterministic model. Such a consideration might have ramifications in the field. As EPA and the registrant continue to explore sampling plans for monitoring, some thought should be given to actually using the simulation models and CWS characterizations as part of the monitoring process. In particular, it is feasible that models will eventually be accurate enough to provide predictions of atrazine concentrations in source waters to a CWS for the coming crop season. Instead of requiring a CWS to collect and analyze water samples in their output stream (drinking water) at some pre-defined frequency (e.g., daily or weekly in the case of some sites), it should be possible to use the models to facilitate targeting sampling to periods of time most likely to experience an exceedance. The Panel suggested a sampling plan that closely follows sampling protocols typically used for obtaining information on rare events. It would control the sampling effort and, at the same time, focus sampling on periods with the highest likelihood of actually seeing measurable concentrations.

Charge Issue 5: Potential Sensitivity of Infants and Children - Consideration of the value of the FQPA safety factor is seemingly best predicated on transparent and systematic consideration of the most important qualitative and quantitative uncertainties associated with both exposure and effect, relevant to susceptible life stages, in a context consistent with that for other pesticides. While considerable information related to water monitoring is available, dietary intake of atrazine in young children or adolescents remains somewhat uncertain and at the present time, it remains unclear what the relevant duration of dosing for either adults or developing organisms may be. There is a general consensus from the results of the animal studies that atrazine can influence the gonadotropin-releasing hormone (GnRH)-mediated LH surge following oral exposures leading to reproductive and/or developmental toxicity. One of the most important uncertainties from available toxicokinetic studies is the magnitude of the internal dose in the adults and the fetuses/pups, but newer studies seem to be pointing in the right direction to characterize these values. An extension of the data on internal dose is needed for developing a PBPK model that will better estimate the internal concentrations in people. This activity seems to be underway and so a PBPK model should be forthcoming. None of the experimental studies done thus far indicate that the prenatally, lactationally, or peripubertally-exposed animal exhibits higher sensitivity to developmental disruption than the effects in adults related to changes in LH surge. Thus,

it remains unclear whether existing data coupled with as yet unfinished studies will answer sufficiently the questions about the toxicity profile for all important life stages (gestational, lactational and peri-pubertal) and whether there are any life stage-related differences in sensitivity. One of the biggest uncertainties at the present time is the validity of the very low dose effects observed on mammary gland development in the Enoch *et al*. (2007) paper which suggests that alterations in development occur following repeated exposures to a mixture of atrazine and metabolites/degradates. It should be stressed that earlier studies also reported mammary gland effects with early in life atrazine exposure (Rayner *et al*., 2005), but with markedly higher exposures. The study reported by Enoch *et al*. (2007) is the only study evaluating mixtures of atrazine and its degradation products. While the mechanism for the non-cancer mammary gland effects following atrazine/degradate exposures is unknown, findings from studies to date, along with the apparently substantial accumulation of atrazine and/or its metabolites in the mammary gland, provide some concern for higher sensitivity in developing organisms. The (re)-consideration of the FQPA 10X Safety Factor should start with documentation and conclusions regarding the adequacy of the data describing atrazine's toxicity profile in adults in order to make intelligent conclusions regarding comparative sensitivities. Once that is completed, the data for earlier life stages can be presented and more clearly interpreted.

Charge Issue 6: Implications of MOA and Toxicity Profile on Water Monitoring - With regard to the water sampling frequency, several Panel members indicated that they believed the sampling frequency currently being conducted by the registrant was adequate. No one raised strong objections or could offer a rationale for an alternative sampling frequency based on the collected information discussed during the SAP meeting. There are some assumptions and extrapolations that contribute to the proposed critical window of human exposure but given the collective uncertainties that these assumptions introduce, the imprecision in the Agency's proposed sampling frequency seems justified. This may be about as precise an estimate as can be obtained when starting with the experimental animal data and the exposure requirements for LH surge suppression as opposed to using outcomes that are more unequivocally adverse. One question that clearly needs further consideration is whether there is a critical exposure (such as a minimum AUC at the target site) that leads to a given level of suppression of the LH surge.

The Panel offered the following approach for setting the boundaries on exposures of concern for human health effects. The Agency has appropriately concluded that the limited epidemiological human evidence is insufficient to establish causality and does not provide sufficient quantitative exposure information to use in a risk assessment. However, what if one assumes that the reported human health outcomes are, in fact, due to current levels of exposure to atrazine? The patterns of atrazine concentrations in water could then be used to provide reasonable estimates of the extent and duration of human consumption of atrazine following agricultural applications for pre-emergent weed control. Simple models could be used to estimate human exposures corresponding to a range of times of exposure to the elevated concentrations observed in the field, given the expected maximum water consumption. Where spikes in water concentration are short

lived and concentrations of atrazine change over a short time scale, then steady state blood levels will not be achieved. In this instance, one might consider what would be average levels of exposure, relate this to the equivalent internal exposure in the rat, and use the promised PBPK model to extrapolate to humans. This may represent a reasonable alternative approach to getting at levels of atrazine in drinking water that may represent risks to human health. These risks could be compared, certainly on an order of magnitude scale against those calculated from the animal data, and may provide a lower bound conservative floor from which to work, and provide a different perspective on the water sampling frequency problem. This would put the Agency in a much better position if, in fact, the Agricultural Health Study or other epidemiology studies provide further support for human health effects as the results continue to accumulate.

The other consideration when faced with the uncertainty over a critical exposure period of from a few days to 4 weeks is whether basing sampling frequency on human health effects is, in fact, the best course of action. It may be more useful to consider a strategy that attempts to capture the pattern of atrazine concentrations in the source water of each CWS based on the characteristics of that particular water system, as opposed to a one-size-fits-all approach based on the series of health-based considerations put forth by the Agency. Given the collective limitations of the health outcome-based approach, this would seem prudent, and would again put the Agency in a better position to take further action should the results of ongoing or future epidemiology studies prove more convincing. In the meantime, since water is the primary source of environmental exposures to atrazine and its metabolites/degradates, there is value in doing a better job of establishing the relationship between the measured concentrations at the community level, and the resulting absorbed dose in humans.

PANEL DELIBERATIONS AND RESPONSE TO CHARGE

As part of the re-evaluation of the health effects of atrazine, three meetings of the FIFRA Scientific Advisory Panel (SAP) were scheduled in 2010. The first two meetings were held in February and April. The purpose of the September meeting was to solicit feedback from the SAP on the status and overall scientific direction of the Agency's re-evaluation on several topics including our preliminary conclusions regarding the non-cancer epidemiology literature on atrazine, as well as the human health risk implications of the experimental toxicology data. The Agency also sought the Panel's feedback on the proposed updates to the dose-response assessment including the use of an internal dose metric and EPA's benchmark dose analysis. With respect to drinking water exposure, the Agency sought the Panel's feedback on a general framework for designing monitoring studies along with basic approaches for analyzing sampling strategies and other methods for estimating exposures from less frequent existing monitoring data. Finally, the SAP was asked to consider and comment on the implications of atrazine's toxicity profile and its MOA on the development of a water monitoring strategy.

Question 1.0: Non-cancer Epidemiology

Section 3.0 and Appendix B of the draft Issue Paper provide the Agency reviews and synthesis of the non-cancer epidemiology studies available for atrazine. These include studies on a variety of topics, notably female and male reproductive outcomes and birth outcomes in addition to other topics. Section 4.0 integrates the findings of the epidemiology and experimental toxicology studies.

Question 1.1

Please comment on the sufficiency of the Agency's non-cancer epidemiology reviews with respect to identifying the major strengths and limitations of each study.

Panel Response

The Panel congratulates EPA on its inclusion and evaluation of the epidemiologic literature as a component of the risk assessment process. The EPA critique of the literature on non-cancer outcomes in relation to potential atrazine exposure has been thoughtful and comprehensive. While it is easy to criticize human studies because, generally, one cannot control exposures and behaviors as well as in laboratory animal studies, human data do have strengths that are important for risk assessment. Many of the problems in measuring exposures and outcomes, especially non-cancer outcomes, can be mitigated with robust study designs.

Some of the general comments and recommendations which follow also were made in the February 2010 Scientific Advisory Panel review but do not appear to have been heeded in the present draft of the Issue Paper. We encourage EPA to incorporate the earlier advice regarding evaluation of epidemiologic studies.

The Panel disagreed with the Agency's conclusion that the database for non-cancer outcomes is strong (i.e., it started with 29 candidate papers and reduced the number to 19 papers; 38 additional papers were cancer-related). Only 19 papers met the Agency's criteria for evaluation and these covered many different health outcomes of interest and had many important limitations, as were noted both by EPA and in the Panel's own detailed analysis presented below. The cancer study database will be larger and more relevant for men. Women are generally under-represented in studies of occupational exposures.

The Panel had a number of comments and recommendations regarding the methodology used for evaluating these papers. First, the criteria for including studies should be described in greater detail. Second, EPA should include an appendix with a full citation and abstract for each paper excluded from full evaluation and a summary describing why the paper was excluded so that EPA can demonstrate the completeness of its review and justify the exclusions. Third, of the 29 papers identified (page 324 of the Issue Paper), two were excluded primarily because they did not independently measure atrazine or triazines. This is not well justified in the text. Other studies that were included, either ecologic or observational in design, did not report any semi-quantitative measures of atrazine or triazine exposure. For example, season of birth or distance to corn or soybean crops may be a surrogate or proxy of exposure, but these are not direct measures of atrazine or triazine exposures. Therefore, neither of these surrogates is likely to be correlated closely with actual exposures or absorbed doses in humans. Fourth, of the studies identified in the literature search, 46 identified as exposure assessment studies were excluded from further evaluation. For the purposes of the epidemiologic review of reproductive and other effects, these are not needed, but in the future (especially for the evaluation of cancer risks), these should be reviewed and summarized by EPA exposure assessors and epidemiologists and it was suggested that the review and summary be provided to the SAP for consideration. When the epidemiologic database is sufficiently strong to support dose-response assessment, such studies can be used to connect the exposure assessment data with the questionnaire and biomarker data, given knowledge of the toxicokinetics of the herbicide.

Fifth, it may be efficient to develop a scoring system for the quality of studies (i.e., how well do they meet the inclusion criteria) to show how each paper addresses the key factors identified by the EPA and the February 2010 SAP Panel (statistical power, sample size, generalizability, accurate and precise exposure and outcome assessment methods, sophistication of statistical analysis, control for confounding, minimized participation and selection biases, etc.). These have been developed for other types of epidemiologic analyses (e.g., meta analysis, Cochrane reviews) and could be applied easily by EPA. Also, in future reviews of other pesticides and for larger epidemiologic databases, EPA may want to have two reviewers independently examine and score each paper (perhaps, initially on quality and then on potential for use in dose-response assessment).

Lastly, it seems that the bar for review of the ecologic studies in the present Issue Paper was set rather low. Ecologic studies depend on aggregate data for exposure and

outcome, and the Panel believed that this study type should be considered in a lower class separate from cohort and case-control studies of individual exposures and outcomes, that they should rarely be used for evaluating weight of evidence, and are useful only for purposes of hypothesis generation. If some reasonable measure or proxy of exposure (e.g., surface, groundwater or finished drinking water contamination levels) for the group level data is available, the studies that include such exposure measures should be given more weight than those that have no measures or proxies (e.g., season of birth).

In general, EPA has done a good job of improving an Issue Paper that was already strong. Four problems are typically common to all major public health issues involving quantitative data. They are the following: 1) the data are vast; 2) the data are highly complex (in disciplines, problems addressed, methods, settings, etc.); 3) the data are generally of poor quality, and 4) the data may just not be what assessors would like to have (e.g., high-dose lifetime animal ingestion studies are conducted when low-dose intermittent human exposures by inhalation or consumption of drinking water would be of more interest). These problems also apply with some force to the data relevant to atrazine and it is apparent that EPA has largely attempted to consider and address these problems in evaluating the literature.

The Panel's evaluation of the relevant epidemiology literature and of the Agency's review of these studies follows.

Female Reproductive System

(1) Farr *et al*. (2004), (pages 30 and 335-6 of the Issue Paper): The EPA critique was largely appropriate in highlighting the strengths and limitations of the Farr *et al*. (2004) analytic approach to evaluating the relationship of atrazine exposure to menstrual cycle characteristics in the Agricultural Health Study (AHS). EPA was correct in noting that lifetime exposure is assessed while menstrual cycle outcomes are captured either within the past 12 months or as average cycle characteristics, which might result in non-differential misclassification of either the relevant exposure or the outcome, and thus would tend to attenuate any true association, if one exists. This approach also does not address the temporal relationship of exposures to outcomes. The Issue Paper might also note that Farr *et al*. (2004) relied on recall of exposures and menstrual cycle outcomes, and, although it was a cohort study, it was analyzed with odds ratios (OR), which might enhance estimates of risk somewhat. Some important aspects of this study were not described in the Issue Paper. First, the participation rate of women included in this analysis, based on data from the AHS, was only 57%, which raises the potential for participation bias (e.g., hypothetically, women with menstrual abnormalities or concerns about pesticide exposure might have been more likely to participate, making the data less representative and potentially overestimating the relationship of atrazine to long menstrual cycles and missed periods, the strongest associations found for atrazine). Secondly, at issue is the representativeness of the study population; the AHS is 97% white and fairly well educated so that the generalizability of the findings from this study is somewhat

limited, and any effect modifiers, such as race/ethnicity or socioeconomic status, cannot be examined effectively. This point was included in the February 2010 SAP comments, but not in the September EPA draft Issue Paper.

(2) Farr *et al*. (2006), (pages 30-31 and 332-3 of the Issue Paper): The EPA critique appropriately described the strengths and limitations of the second paper by Farr *et al*. (2006)on the AHS. This paper examined the relationship of exposure to age at menopause, an indicator of subsequent disease risk and longevity (i.e., later age at menopause is associated with reduced risks to, and mortality from, a number of chronic diseases and, therefore, with increased longevity, although it is associated also with an increased risk of breast cancer). However, the ninth line on page 31 is incorrect; a hazard ratio of 0.79 for atrazine exposure does not constitute a "greater" association than that for use of hormonally active or ovotoxic pesticides, which was 0.77, because the closer the hazard ratio is to 1.0 (regardless of whether the hazard ratio is less than or greater than 1.0), the smaller the association. The EPA critique is correct in pointing out that a non-standard question was used to assess menopause and peri-menopause, that pesticide exposure was assessed, rather than, specifically, atrazine exposure, and that timing of exposure relative to the outcome could not readily be addressed in this study. The Issue Paper should mention that the initial participation rate in this portion of the AHS was relatively low (59.5%), thus again creating the potential for participation bias, albeit probably not great for this outcome. Also, as before, the generalizability of the results from the AHS is somewhat limited due to the lack of heterogeneity in race/ethnicity, socioeconomic characteristics and lifestyle of the study sample.

(3) Saldana *et al*. (2007), (pages 31 and 354-5 of the Issue Paper): The EPA critique is appropriate in the analysis of the AHS by Saldana *et al*. (2007) on the relation of pesticide exposure to gestational diabetes. EPA might add to the next version of the Issue Paper the point that while the relationship of atrazine exposure to gestational diabetes is depicted only graphically, the adjusted odds ratio is in excess of 1.0, and the 95% confidence limits do not appear to include 1.0, which differs from the indication on page 354 that the association was not statistically significant. Limitations of the generalizability of the results of the AHS due to inclusion of a primarily white and relatively homogeneous study sample, as noted above, apply to this paper as well.

In summary, these studies evaluating the associations of atrazine with female reproductive outcomes (longer menstrual cycles, more missed periods, later age at menopause, and gestational diabetes) are suggestive, but inconclusive because of the lack of representativeness of the study sample, the potential for participation bias, non-concordance of the periods of exposure with those of the outcomes, and use of non-standard questions for some outcomes. Also, the use of odds ratios may overestimate relative risks and, thus, the magnitude of association. Some of these issues regarding the limitations of the AHS were raised in the February 2010 SAP's report on the use of epidemiologic studies in risk assessment, but appear not to have been incorporated in the

critiques included in this Issue Paper. If EPA disagrees with the Panel's comments, the next version of the Issue Paper should cite the reasons.

Furthermore, while the AHS included a study sample with likely substantial pesticide exposure, thus, representing a potentially high risk population, it was not designed to address questions related to reproductive outcomes. Other epidemiologic studies have been better designed to assess menstrual cycle characteristics and age at menopause, have collected biologic samples that could be assessed for atrazine exposure, and/or have better representativeness and diversity in their study samples than the AHS. Examples of such studies include the Study of Women's Health Across the Nation (SWAN) (Gold *et al*., 2001; Sowers *et al*., 2000) and perhaps the Midlife Women's Health Study (Bromberger *et al*., 1997) or the Australian longitudinal study of midlife women (Burger et al., 1995), the Semiconductor Women's Health Study (Gold *et al*., 1995), and/or the Sacramento Community Health Study (Gold *et al*., 2010). Evaluating data and samples from these other studies would make the findings more generalizable, and would permit examination of potentially important modifications of the effects by race/ethnicity, lifestyle or health-related factors, which are important in gene-environment interactions and potential epigenetic effects. Additionally, the concern expressed in the EPA critique regarding adequate control of physical activity is overstated for age at menopause as few, if any, studies have shown this to be an important determinant of age at menopause; key factors for this outcome are smoking, educational level and use of oral contraceptives. Finally, in several reports, atrazine was studied along with other pesticides. Results for atrazine then are presented in comparison with the results for all pesticides. This is incorrect; the result of such an analysis is to reduce the apparent strength of evidence for some effect. The comparison should be between atrazine and all other pesticides in the study or to no pesticide exposure. It may be that the original authors often or always reported the wrong comparisons in a way that cannot be corrected, but EPA should at least note the error.

Male Reproductive System

> **(1) Swan *et al*. (2003), (pages 32-3 and 362-4 of the Issue Paper):** The EPA critique of the strengths and limitations of the Swan *et al*. (2003) cross-sectional case-control study is appropriate. Like the AHS, the study sample, derived from the multicenter Study of Future Families, was almost entirely white, once again limiting the generalizability of the findings and the ability to examine interactions. In addition, the participation rate was low (27%), and the sample size was small; however, the fact that the study was hypothesis-driven and biomarkers of exposure and outcomes were used were important strengths. A small correction should be made in the second line of the critique on page 362 to indicate that this study was of non-persistent pesticides (rather than persistent, as currently indicated on this page). The study authors found that high atrazine levels were associated with poor semen concentration, morphology and motility.

This single study examining the relationship of atrazine exposure to reduced sperm quality was suggestive with the important strengths of being hypothesis-driven and

using biomarkers of exposures and outcomes, but its limitations need to be addressed in additional studies.

Fetal and Perinatal Outcomes

(1) Arbuckle *et al*. (2001), (pages 33 and 327-8 of the Issue Paper): The EPA critique of Arbuckle *et al*. (2001) is mostly appropriate with the exception of the exposure assessment issue described below. The Issue Paper should note that the Ontario Farm Family study used a historical cohort design. Data were analyzed with odds ratios, which may overestimate relative risk, without adjusting for confounders. Some analyses appeared to use post-conception atrazine or other herbicide exposure or later preconception exposure as the referent category, presumably to try to address potential recall bias, but the number of women who reported spontaneous abortions and were exposed in this latter category was small. This publication also was unclear about how multiple pregnancies from the same woman, which represent lack of statistical independence, were handled statistically. (Tests for non-independent outcomes exist, but were not used here.) The Issue Paper also should emphasize that the associations of preconception atrazine exposure with early spontaneous abortions were modest in magnitude and not statistically significant (ORs ranged from 1.3-1.7, depending on the referent used, all with 95% confidence intervals including 1.0). In addition, the last paragraph of the Agency's review of this paper should refer to "odds" not "risks" because odds ratios and not relative risks were determined. The EPA critique of this and the Savitz *et al*. (1997) paper seems to discount results because they used farm level exposures rather than direct personal exposure measurements (page 328) so that farm level exposures of male and female applicators cannot be distinguished. This approach, however, might represent a much better effort at historical pesticide exposure reconstruction than the self-reported measures used in other observational studies (e.g., the AHS ever/never classification). The Issue Paper correctly notes that women who were aged 35 years or older and reported preconception exposure to triazines had a nearly three-fold, statistically significantly, increased odds of spontaneous abortion (OR 2.7, 95% CI 1.1, 6.9). Women aged 35 years or older who reported preconception use of both triazines and thiocarbamates had an OR=7.5 for spontaneous abortion, all of which suggests that older maternal age may modify the overall risk of preconception triazine use by increasing the odds of spontaneous abortion.

(2) Savitz *et al*. (1997), (pages 34, 39and 358-9 of the Issue Paper): The EPA critique of the historical study by Savitz *et al*. (1997) is appropriate. In addition to not being able to link farm activities with the application of crop herbicides and the use of specific herbicides (as noted in the critique), the small sample size of women exposed to atrazine who had some of the outcomes (e.g., preterm and small-for-gestational-age deliveries) and the modest participation rate (64%) also are limitations, although they probably resulted in less likelihood of detecting differences as statistically significant. Paternal use of atrazine on crops had an

adjusted odds ratio for miscarriage of 1.5 (95% CI 0.9-2.4), for preterm delivery of 2.4 (95% CI 0.8-7.0) and for small-for-gestational-age of 0.5 (95% CI 0.2-1.3).

In summary, these two studies of the relationship of atrazine exposure to fetal and perinatal outcomes were suggestive, with largely positive but non-significant associations that were modest in magnitude. They also possessed a number of limitations, including use of odds ratios instead of relative risks and small sample sizes for some outcomes, resulting in wide confidence intervals, lack of adequate statistical power, lack of clarity in the handling of lack of independence of multiple pregnancies from the same woman, lack of data on specific herbicide exposure, and potential for participation bias. These deficits should be addressed in additional studies.

Birth Defects

(1) Mattix *et al*. (2007), (pages 34-5 and 344-5 of the Issue Paper): The EPA critique of the paper by Mattix *et al*. (2007) correctly notes its major limitation: the ecologic study design which does not permit assessment of actual maternal exposure to atrazine in the drinking water. Thus, the correlation between atrazine concentrations in surface water and month of conception of fetuses with abdominal wall defects does not mean that the individual mothers who delivered such infants actually consumed water with elevated atrazine.

(2) Hornemann *et al*. (2009) (pages 35 and 340 of the Issue Paper): The EPA critique appropriately notes the limitations of this ecologic study by Hornemann *et al*. (2009). These authors examined seasonality of omphalocele prevalence rate during a time when atrazine was banned from use in agriculture in Germany so that they had little likelihood of observing differences in prevalence rates by season. Also, the ecologic design would not permit assessment of actual individual maternal prenatal atrazine exposure in relation to occurrence of omphalocele.

(3) Winchester *et al*. (2009) (pages 35-7 and 370-1 of the Issue Paper): The EPA critique appropriately indicates that the major limitation of the Winchester *et al*. (2009) study is its ecologic design, which does not permit assessment of the individual maternal exposures to atrazine or other contaminants consumed in drinking water in relation to occurrence of birth defects. The critique also appropriately notes that the effect sizes are small.

(4) Ochoa-Acuna and Carbajo (2009a) (pages 36 and 352 of the Issue Paper): The EPA critique correctly identifies the major limitation of the Ochoa-Acuna (2009) study as being the use of residence near corn fields as a surrogate of potential chemical exposure, so that actual individual maternal exposures were not measured. In addition, the sample sizes were small for several types of birth defects, and it was likely that chance could have resulted in at least some of the significant findings due to the multiple statistical tests that were performed. Also, while the authors indicated that atrazine is used largely on corn fields, rather than

soybean fields, other chemicals such as fertilizer nutrients also are preferentially applied to corn fields. The study authors did not indicate whether other pesticides also were used on corn fields, which might also be a potential explanation for the observed findings.

(5): Waller *et al*. (2010) (pages 36-7 and 368-9 of the Issue Paper): The EPA critique of the population-based case-control study of gastroschisis reported by Waller *et al*. (2010) is quite appropriate in noting the strengths of the case-control approach, including individual information on exposures and outcome, compared with that of an ecologic design, and the large sample size studied, while also noting the limitations, including: a) that the exposure was largely ecologic (i.e., distances of residence to "high" atrazine surface water monitoring site, rather than actual maternal tap water exposure); b) the lack of analysis of surface water atrazine concentration by maternal age; c) the possibility of misclassification of exposure due to monitoring of sites that were not necessarily the source of drinking water and the possibility of mothers relocating residence between early pregnancy and delivery; and d) the lack of clarity of selection of confounders to control in multiple logistic regression models. In addition to the variables already listed in the Issue Paper, marital status, race and income also should be noted as differing significantly between cases and controls but not controlled in the multivariable analyses. The study authors concluded that the occurrence of gastroschisis was significantly inversely related to distance of the maternal residence from the closest "high" atrazine monitoring site.

In summary, three of the five studies of birth defects used the sub-optimal ecologic design so that atrazine exposures and outcomes were not assessed on an individual level. This is a major limitation. The other two studies are also considered only suggestive because one did not examine individual maternal exposures to atrazine and the other used a largely ecologic measure, rather than an individual measure of atrazine exposure. Additional limitations included: multiple statistical testing that might have resulted in statistical significance due to chance alone because the p values were not adjusted for multiple testing; the analyses did not account for other exposures that might explain the observed effect (Ochoa-Acuna); and confounding variables were not adequately described and adjusted for in analyses (Waller).

Adverse Birth Outcomes

(1) Ochoa-Acuna *et al*. (2009b) (pages 38 and 349-50 of the Issue Paper): The EPA critique of the population-based, historical cohort study by Ochoa-Acuna *et al*. (2009) on the relationship of atrazine concentrations in drinking water to the prevalence of pre-term and small-for-gestational-age infants appropriately noted the strengths of using a more refined exposure measure of atrazine based on monitoring data and adjustment for multiple individual-level confounding variables. The Issue Paper, however, also should note the strength of the large sample size, as well as the limitation that individual maternal tap water consumption was not obtained, which might have resulted in some

misclassification of exposure. No association was found between atrazine concentrations and preterm births, but a significant positive association (a 20 % increase) was found with small-for-gestational-age.

(2) Villanueva *et al*. (2005) (pages 38-40 and 365-6 of the Issue Paper): The EPA critique of the Villanueva cross-sectional study appropriately noted its strength in using measurements of atrazine in drinking water and the major limitation of uncertainty in using these measurements to estimate individual maternal atrazine exposures as well as the large number of measurements below the limit of detection and the resulting narrow range of exposure levels, which reduced the ability to detect significant differences in exposure groups. The study's sample size was moderate for the three outcomes of interest: low birth weight, preterm delivery and small-for-gestational-age. Indeed, no significant associations were detected between atrazine concentrations in drinking water and these three pregnancy outcomes. The Issue Paper also should note that the findings are suggestive of elevated odds of low birth weight (which was not statistically significant) and small-for-gestational-age in pregnancies whose third trimester overlapped high atrazine use periods (50% increase) and, similarly, for preterm deliveries whose first trimester overlapped these periods (not statistically significant). Importantly, it also should be noted that the analyses in this paper were not adjusted for variables that are often influential in these outcomes, such as maternal smoking, educational level and parity.

(3) Munger *et al*. (1997) (pages 38-9 and 347-8 of the Issue Paper): The EPA critique of the ecologic study by Munger *et al*. (1997) is quite appropriate in noting that the correlation of drinking water levels of atrazine and other triazine herbicides with low birth weight, prematurity, and intrauterine growth retardation is useful only for generating hypotheses to be studied in methodologically stronger epidemiologic approaches because time periods for the exposure levels and birth outcomes were not completely comparable. The analyses assume that the relative ranking of the communities with regard to drinking water contaminants remained the same throughout the study period. Again, the study population was largely white (thus limiting the generalizability of results). The Issue Paper should note that this study did not examine individual maternal drinking water exposures and controlled only for some of the important confounding variables (e.g., smoking and maternal educational level) on a community rather than individual level, thus potentially misclassifying or inadequately controlling for confounders on an individual level.

(4) Dabrowski *et al*. (2003) (pages 38-9 and 330 of the Issue Paper): The EPA critique of the population-based, case-control study by Dabrowski *et al*. (2003) of the relationship of pesticide exposure during pregnancy to birth weight and pregnancy duration appropriately notes that the strengths of the study included its population-based comparison with a control group similar to the cases, ascertainment of birth weight from medical records, paternal confirmation of pesticide exposure reported by mothers and adjustment for multiple confounding

variables. The critique also highlights the major limitation that the analysis is not specific for atrazine exposure, and the sample size was too small to examine specific pesticides. The results indicated a significant reduction in birth weight and pregnancy duration with pesticide exposure in the first or second trimester after adjustment for important confounders.

(5) Sathyanarayana *et al*. (2010) (pages 39 and 356 of the Issue paper): The Sathyanarayana *et al*. (2010) study analyzed the AHS for the association of maternal pesticide exposure to birth weight appropriately. EPA noted that the study did not specifically address atrazine exposure. Additionally, the number of low birth weight or preterm infants was too small to assess individual pesticide exposures or to detect meaningful differences while adjusting for multiple confounding factors. Indeed, the study found no significant effect of maternal pesticide exposure, even direct agricultural exposure, on birth weight. A strength of the study was the control for multiple confounding variables in the analyses. As noted previously, the AHS is comprised of a largely white population sample, which limits the generalizability of the findings and the ability to examine any modification of effects by race/ethnicity, socioeconomic status or lifestyle or health factors.

In summary, these five studies of adverse pregnancy outcomes ranged from ecologic to population-based case-control and historical cohort designs, the latter having the strengths of better individual assessment of exposure to atrazine. Strengths included some large sample sizes and adjustments for multiple confounding variables to assess the independent effect of atrazine exposure, although individual maternal water consumption was not assessed to improve estimates of individual exposures. The findings suggest a relation of atrazine exposure to small-for-gestational-age, but two of the non-ecologic study designs focusing on birth weight were not specific for atrazine. Small-for-gestational-age has many causes, and more than one cause (if any) may be affected by a pesticide. Lack of agreement among studies may be a result of multiple mechanisms. For example, atrazine may have a direct effect on the fetus at some time or in some settings and work through the mother's endocrine system in others.

Respiratory Effects

(1) Hoppin *et al*. (2002) (pages 40 and 338-9 of the Issue Paper): The EPA critique appropriately noted that this paper reporting on a cross-sectional association of wheeze with pesticide use in the AHS showed a modest but significant association of atrazine with wheeze (odds ratio =1.2, 95% CI 1.07, 1.34) with a significant dose-response trend observed with days of use, although this was not an *a priori* exposure hypothesis. As noted in the critique, this study did have a large sample size and was able to control for multiple confounding factors, especially exposure to corn and grain dust. The Issue Paper also should note that this study had the same limitations noted previously for the AHS, particularly the potential for participation bias and the lack of diversity in the study sample which reduces generalizability and the ability to examine

modification of the effects by race/ethnicity, socioeconomic status and lifestyle or health factors. As noted in the Issue Paper, a subsequent paper with overlapping authorship examined the same study sample and found that stratifying on two different conditions (allergic and non-allergic asthma) and controlling for additional potential confounding factors resulted in no significant relationship of atrazine exposure to these outcomes (Hoppin *et al*., 2009).

(2) Kossman *et al*. (1997) (pages 40-1 and 342-3 of the Issue Paper): The EPA critique correctly identifies the limitations of the cross-sectional study by Kossman *et al*. (1997), including the inability to evaluate the specific relationship of atrazine exposure to impaired respiratory function and the lack of control of confounding factors (particularly importantly, smoking). In addition, the Issue Paper should note that the sample size was small and would not have provided adequate statistical power to detect meaningful associations if specific chemicals were evaluated and if the associations were controlled for confounders.

(3) Kluchinski *et al*. (2001) (pages 41 and 341 of the Issue Paper): The EPA critique correctly identified the limitations of this cross-sectional study to include the inability to evaluate the specific relationship of atrazine exposure to immune parameters that may be biomarkers of impaired respiratory function and the lack of control of confounding factors (particularly smoking). In addition, the Issue Paper should note that the sample size was small and would not have provided adequate statistical power to detect meaningful associations if specific chemicals were evaluated and if the associations were controlled for confounders.

In summary, these three cross-sectional studies of the relation of pesticides to respiratory impairment had significant limitations, including sample sizes inadequate to provide sufficient statistical power to detect meaningful effects associated with specific chemicals while controlling for confounding variables; any observed associations were modest in magnitude.

It appears that relatively little still is known about human fertility related to atrazine exposure. This reality should be noted in the Issue Paper. If additional epidemiologic studies are undertaken, they should be well-designed with sample sizes sufficient to identify meaningful effects while controlling for confounding factors and permitting examination of effect modification by important genetic/demographic/ lifestyle/health factors, adjustment for multiple comparisons, standard and precise assessment of individuals' actual exposure and outcomes and more representative and diverse study samples. New research should focus on relevant health outcomes in individuals who have relevant, real-life exposures (e.g., in occupational settings, such as farm or manufacturing workers or individuals with likely or known drinking water exposures), perhaps in appropriate agricultural areas. Such studies will need to be supported by laboratory data bearing on important public health issues so that significant effects or observations of no significant association(s) in humans can be appropriately interpreted.

One Panelist expressed concern about the handling of studies with substantial numbers of analytical chemistry results below the limit of detection (LOD). A lot of useful information can be derived from such findings, even when they account for most of the observations, and especially when interest is focused primarily on the highest exposures, as is the case in many of the studies presented in the Issue Paper. At the very least, one can do a rough test by first assuming that all of the "below LOD" values are, in fact, zero, then assuming that all are at the upper end of the no-detect range. Regardless of whether these two analyses are in rough agreement on the most important conclusions, the Issue Paper should say so. More sophisticated ways are available to approach the LOD issue.

Question 1.2

At this time, atrazine's non-cancer epidemiologic database is not robust enough for inclusion in quantitative risk assessment and does not support the ability to determine causal associations. There are several limitations present in the current database, and the quality of atrazine exposure assessment is paramount among the limitations. In Section 4.0 of the draft Issue Paper, the Agency describes the qualitative similarities and differences between the epidemiologic findings and the experimental toxicology database. In short, the observational studies – particularly those related to reproductive effects in adult females, as well as small for gestational age in newborns – lend further support for the human relevance of the laboratory animal findings.

a. Please comment on the scientific information that does and does not support the Agency's conclusions (as described in Section 3.0) with respect to the characterization of quality, and limitations of the non-cancer epidemiologic database and its utility in hazard characterization, dose response analysis, and quantitative risk assessment.

Panel Response

EPA states (Section 4, pg 58) that the epidemiology and toxicology databases can be used together to inform hazard identification, dose-response assessment and hazard characterization, but for purposes of quantitative risk assessment (i.e., risk characterization), the experimental toxicology data (particularly, *in vivo* data from the rat) will be used for endpoint selection and dose-response assessment.

EPA has conducted a thorough study-by-study review of the non-cancer epidemiologic papers; the SAP critique of the Agency review is presented in the response to Question 1.1. Overall, the SAP agrees with the Agency's conclusions that the current non-cancer epidemiologic database (n=19 studies) provides some useful information for hazard identification but is too limited to provide sufficient information for credible dose response assessment or risk characterization. As the Agency states, exposure characterization is the primary limitation; it would be impossible to reconstruct individual level exposures using the available observational data. Furthermore, many of the 19 studies are ecologic in design and should be given little weight for risk assessment purposes.

Some general comments on the broader issues regarding interpretation of the integrated data were presented and deserve more explicit acknowledgement and discussion in the EPA draft. These extend across a whole set of studies and are not limited to the interpretation of single studies. First, a general failing in most studies of toxic effects is the lack of sufficient attention to matters of statistical power. "Lack of evidence showing an effect" is simply not the same as "evidence showing a lack of effect." The EPA draft would be stronger if it discussed the likelihood that critical (apparently) negative results might be reversed by better and stronger designs and, especially, larger studies (i.e. with greater statistical power), whether in humans or animals and whether findings, even when non-significant, show consistency across studies with different designs performed in different populations.

Similarly, the "multiple comparison problem" is not sufficiently appreciated. This problem is widely recognized as it applies to single studies; but it applies equally to any collection of papers or other sets of independent results. Many studies have examined the health effects of atrazine, each with multiple endpoints accompanied by many analyses of their data. At the traditional default 5% level for statistical significance in scientific analyses, an average of one in 20 tests, in which no effect truly exists, will be labeled "statistically significant." This 5% approach (or 95%, in the case of confidence bounds) has been very productive for many decades across all areas of science, as important results can be evaluated in a broad context of prior knowledge and, sometimes, by the collection of new data. The multiple comparisons problem is pervasive and hard to deal with in reviews that cover large numbers of reports. The question in such reviews is not whether some result is statistically significant, but whether the overall pattern of results labeled significant actually lies outside the expected range. Roughly speaking, if the literature reported 100 tests of data in which no true effect exists, one should expect about 5 to be "significant." If say, 10 are significant, or if some of the reported p-values are quite small, an effect may be present. The SAP urges EPA to review the critical positive findings and to comment at greater length on the plausibility that such findings are a result of multiple comparisons.

One more point is that statistical confidence bounds and p-values measure the effects of random error, and random error only. Bias increases the uncertainty, so that calculated p-values are too small and confidence bounds are too narrow, and the effect estimate is generally not accurate. Bias can be made rather small in designed laboratory studies, but is a major threat to conclusions in other settings unless sufficient attention is given to minimizing selection and participation biases in the design and implementation of the study.

These points can be summarized to say that negative findings do not always mean no effect is present, and positive findings do not always mean that such an effect is truly operating. This is as true for laboratory studies as it is for epidemiology. Educated scientific judgment cannot be replaced by reliance on mechanical rules about what is a real effect. It would be helpful to have more explicit acknowledgment of points for which judgment (informed by statistical analysis) still has an important role.

The Panel also presented some specific comments on the epidemiology portion of the Issue Paper. The Issue Paper indicates that no national birth defect registry exists in the US (page 37, line 18). In fact, CDC has operated a very broad-based registry for many years which could be used in geographic studies of atrazine. EPA has presented the data for gastroschisis and omphalocele separately; this is appropriate because while they are both abdominal wall defects (AWDs), they are entirely different conditions and almost certainly brought about by different mechanisms. Finally, while exposures may be of short duration, the effects might be expressed over much longer times (e.g., for outcomes such as cancer, age at menopause and menstrual disturbances and possible hormonal disturbances and even birth defects). Thus, an exclusive focus on exposures occurring based only on the first trimester of gestation is not recommended.

b. Please comment on scientific information that does and does not support the integrative analysis and conclusions, contained in Section 4.0, with respect to the similarities, differences, and uncertainties of the experimental toxicology and epidemiologic findings.

Panel Response

In the February 2010 SAP meeting on the development of a framework for incorporating epidemiologic data into the risk assessment process, two case studies were presented (diazinon incident data and atrazine epidemiologic studies), and a draft framework was developed and presented by the EPA. This was reviewed by the SAP at that time and recommendations for improvements to the framework were provided. The work presented by the EPA at the September 2010 atrazine meeting and summarized in Appendices B1 and B2 reflects consideration of the majority, but not all, of these comments/recommendations.

Well-designed observational studies in humans that test hypotheses which are developed, in part, based upon the adverse effects observed in animals and an understanding of the toxicological mode(s) of action are needed to be able to incorporate the epidemiologic data into the risk assessment process and integrate it with the toxicology database. The comments from the Panel members included in this report reflect their assessment of the literature and associated materials that were provided for the September 14-17 meeting and, to a lesser degree, the additional data/studies that were presented in the public comment session on September 15, 2010.

To discuss the integrative analysis of the epidemiology and toxicology literature, the Panel assumed that atrazine or a mixture of parent and/or its relevant metabolites, following repeated exposures, attenuates the LH surge in animals, and when this attenuation is significant, this can result in adverse reproductive effects in animals. Furthermore, the Panel assumed that the mode of action (MOA) is relevant to humans, although some evidence to the contrary was presented in the public comment session (re: hypothalamus GnRH).

While some major uncertainties in the toxicology database (which includes hundreds of studies) were identified in the SAP meeting, atrazine exposure disrupts the regulation of pituitary luteinizing hormone secretion and can alter a number of reproductive functions in rats based on the experimental toxicology. Attenuation of LH may be a precursor to a number of adverse events (depending, in part, upon life stage of the exposed individual), and might lead to altered reproductive function. The available data did not allow a clear consensus to emerge on the necessary percent attenuation needed to yield adverse developmental and reproductive effects. The form of the dose, route of dosing, and frequency of administration all impact on the size of dose that will produce a given defined response. Multiple lower dose exposures over time (as seen in the Cooper *et al*., 2010, studies of 4 day exposures in rats) were more effective for LH surge attenuation than a single high dose (Cooper *et al*., 2000). Repeated atrazine exposures at doses as low as 3.65 mg/kg/day led to disruption of estrous cyclicity and early reproductive senescence. Greater than 80% attenuation of the LH surge in any given 4-day estrous cycle would be needed to observe deleterious effects in the reproductive system effects in rats. There was significant discussion of the effect of the route of administration on the pharmacokinetic behaviour, internal exposure and magnitude of the attenuation of the LH surge or other relevant precursor events. It was noted that the same applied dose had different effects when administered by feeding and oral gavage. A dose by the latter route that produced reproductive toxicity was without effect when administered in the food (according to data presented by the registrant in its public comments). This may reflect different pharmacokinetic profiles being produced by the different methods of administration. The importance of the form of the dose is well known and had been discussed at an earlier SAP (page 14 of the report of the SAP, April 26 - 29, 2010). Work is ongoing to resolve these problems, but clearly there is a need for data that are relevant to human exposure via drinking water. Moreover, some of the toxicologists speculated that the rodent may not be a good model for the preovulatory LH surge in humans. All of these factors are important for the interpretation and design of both the animal and epidemiologic studies. Assuming a relevant MOA, it will be necessary to consider route and duration of exposure, as well as magnitude of exposure/dose, to assess the potential for risks of adverse reproductive outcomes in humans.

Furthermore, some humans may be exposed to atrazine and/or its metabolites over a significant portion of their lifespan; these exposures are not consistent within or between life stages and may be predominantly oral/dietary (unless the population is occupationally exposed, which may represent predominantly dermal exposure) and are also likely to occur in mixtures. In most animal studies, exposures are mainly to single compounds, generally at higher doses than those to which humans would be exposed. In addition, real-world human exposures may differ in route of administration than those used in the animal studies. Lastly, few mixtures of triazines alone or in combinations with other environmentally-detected substances are generally tested in animal experiments.

As EPA noted, in order to integrate the epidemiology and experimental toxicology information, a strong experimental toxicology database, an understanding of toxicological mode(s) of action, and a strong epidemiology database is needed. For

atrazine, a robust experimental toxicology database exists, including hundreds of studies, and understanding of its likely mode(s) of action is increasing. To complement this toxicology database, we need human epidemiologic studies which have been specifically designed to evaluate relevant outcomes and that are based on defined mode(s) of action. Significant improvements in the epidemiologic literature will be necessary for the integration of these results into quantitative risk assessment.

Since some of the reproductive effects or their precursor events (e.g., estrous cycle characteristics, long or missed cycles, and attenuation of the LH surge) observed in animals are acute in nature and occur after relatively short periods of exposure, prospective studies can be designed to evaluate these outcomes in humans. These can be conducted at relatively low cost (as compared with chronic outcomes such as cancer). To achieve exposures/doses that are measurable and relevant (meaning that exposures are repeated over time, not simply a single occurrence), these types of studies could be conducted in occupational cohorts, such as farm or manufacturing workers, or alternatively, in women living on farms who are likely to have higher exposures than the general population (whose primary source of exposure is via drinking water). Ideally, one would obtain information on current (external) exposures (via direct measurement, questionnaire, records, and/or diaries) and at the same time, collect information on absorbed dose over time (biological samples including urine analyzed for atrazine parent and active metabolites as well as inactive conjugates). This type of information would be far more useful for risk assessment than the measures extant in the epidemiologic atrazine literature to date.

Currently, the epidemiology database consists of 19 non-cancer studies which met EPA's selection criteria for detailed review. EPA states that this is a larger database than is typically available, but these studies cover a broad variety of non-cancer outcomes (ranging from reproductive effects to wheeze) and are not of sufficient quality to include in a comprehensive risk assessment. These 19 epidemiology studies present very little information collected in humans that support the outcomes observed in animals; therefore, the Panel concluded none of these studies were useful for hazard, risk or dose-response assessment and are likely useful only for problem formulation. Epidemiologic data relevant for female or male reproductive health may be obtained from existing cohort studies, which were designed for assessing these outcomes in relation to other exposures or factors and could perhaps be used in hypothesis-generation, but rarely should it be expected that the data will be of sufficient quality to be useful for a risk assessment of outcomes that differ from those of the primary hypotheses. For example, many of the primary outcomes of interest in the Agricultural Health Study (AHS) are cancers and, thus, the methods of exposure assessment were designed to be relevant for cancer outcomes (i.e., long latent periods, some consideration for frequency (seasonal) and duration of exposure (years) and a few select factors that may modify exposures, such as protective clothing worn and equipment used but not genetic, lifestyle or other health-related factors). Furthermore, the AHS was not designed to and thus did not measure well other outcomes (such as menstrual or reproductive outcomes) nor was its study population adequately diverse and large to be more generalizable and to examine effect modification adequately as noted in the previous sentence.

The two studies published by Farr *et al*. in 2004 and 2006, which examined menstrual cycle characteristics and delayed menopause, made use of the AHS cohort but are severely limited by the availability of relevant exposure data. The SAP does not consider that dichotomous exposure categories, which are classified as "ever/never" and the corresponding results or odds ratios will be useful other than for hypothesis generation. They provide no information on the timing, frequency, duration or magnitude of dose – all of these measures are important for the integration of the epidemiologic data with the toxicology data for risk assessment. Although the AHS will provide very important information related to pesticide and other agricultural exposures over the years, particularly in relation to cancer and other chronic diseases, it was not designed to identify accurately these specific reproductive outcomes. The critique of the exposure assessment methods is provided knowing that the study was not designed to test hypotheses about pesticide or atrazine-specific exposure in relation to these reproductive outcomes and thus is not appropriate for assessing risk for these outcomes.

An example of an inappropriate "ever/never" classification of an exposure could be based on the following question: "Have you ever smoked one cigarette?" A large percentage of the population would answer "yes" to this question and would be classified as exposed. This measure would have little relevance for most health outcomes. The more relevant question for a cancer outcome would be "Have you smoked at least 10 cigarettes a day for 10 or more years?" For short-term effects, the question "Have you smoked 10 or more cigarettes/day over the past month?" might be asked. Thus, the hypothesis or study question will dictate the type of questions being asked or the exposure measures used. Furthermore, the use of exposure categories such as "all pesticides" or "herbicides," "insecticides," "fungicides," is not particularly useful. Analyses based on individual chemicals, relevant mixtures, or chemical classes are far more relevant (e.g., triazines as a group) and categorization based on suspected mode of action (i.e., targets a specific gene or endocrine modulators, although this is very broad as well) at least have some biological basis for hypothesis generation.

Of the other studies that indicate some adverse effects in humans, EPA notes the Arbuckle *et al*. (2001, spontaneous abortion), Waller *et al*. (2010, AWDs) and Villanueva *et al*. (2005, preterm delivery, small-for-gestational-age, low birth weight) studies conducted hypothesis testing (Table 3.1, EPA Issue Paper), and two others had non-significant results. Unfortunately, inadequate exposure assessment was a major limitation in these studies. The other studies reviewed, including the ecological studies of group data, should be given little consideration in EPA's analysis. In addition, it should be very clear in the EPA reviews of individual studies that the important distinction be made between ecological studies and the "ecological approach" for exposure assessment. For example, assigning an area value (i.e., average or median concentration of atrazine in finished drinking water from a community water system) to an individual study participant is not an ecological study design. That would represent an observational study with individual assignment of exposure. As noted previously, ecologic studies should be considered only for hypothesis generation.

The Panel agreed that the integration of study findings, interpretations and conclusions in Section 4 of the Issue Paper is largely correct. However, Section 4 does not adequately address the significant limitations of the AHS that were noted directly above and in the response to Question 1.1. Given these limitations and the likely different mechanisms whereby ovarian aging occurs in rodents compared to humans, it may be expected that the experimental toxicology and the limited epidemiologic findings would differ with regard to these effects. Furthermore, as stated previously, other human studies of menstrual cycle characteristics and/or age at menopause are likely to have been better designed to assess these outcomes with less misclassification than the AHS, and many of these other studies have stored biologic specimens that could be measured for atrazine, its metabolites and other environmental contaminants and would do a better job that the AHS of assessing any association of atrazine exposure with these outcomes.

While the limitations of the Swan *et al*. (2003) study have been noted above in the response to Question 1.1, and are important to consider in interpreting the findings, it is also important to note that the findings are congruent with some of the non-rodent findings that were presented at the meeting regarding adverse male reproductive effects. While the participation rate in the Swan *et al*. (2003) study was quite low at 27%, as the authors note, this is not an infrequent occurrence in human studies of semen characteristics and was unlikely to be related to the findings. Furthermore, while the numbers were small, resulting in wide confidence intervals, the effect estimate was large, and the confidence intervals did not include 1.0. It is true that further investigation is needed on the mode of action.

The human data on small-for-gestational-age, low birth weight and prematurity are suggestive and not inconsistent with some of the animal study findings. The Issue Paper appropriately points out the possibility of misclassification, particularly of exposure and/or possibly, of outcome in the human studies. It should be noted that on the fifth line on page 58 the first use of "atrazine" should be "pesticide." It should also be noted at the end of this same paragraph that one would not necessarily expect an effect on birth weight to result from exposure only in the first trimester of pregnancy because an effect resulting from exposure in the third trimester, when fetal weight increase is generally greatest, is at least equally plausible.

An integrative analysis of the epidemiology and toxicology databases could be strengthened with additional information on the toxicokinetics and toxicodynamics in humans or an effective PBPK model, which is expected by 2011. The registrant presented information in the public comment session that humans primarily form the DEA metabolites, while rats primarily form the DIA metabolites. Similar formation of the DACT metabolites is expected. This information is helpful for dose reconstruction in humans based on biological monitoring studies. In addition, a thorough review of the human exposure data (including biological monitoring studies) in both occupational and environmental settings is absolutely necessary to be able to connect potential exposures and absorbed dose measures in humans with doses administered to animals. Such an analysis should be brought to the SAP for peer review.

The EPA conclusions in Section 4 of the Issue Paper indicate that "the non-cancer epidemiology database provides important information that adds to the human relevance of the animal findings, particularly related to female and male reproductive effects and small-for-gestational-age birth outcome." The SAP suggests that only "limited information" is provided. Furthermore, the following statement appears in the EPA Issue Paper: "Among the key studies in the epidemiology database, studies in which the potential for other non-causal explanations is comparatively lower, evidence indicates a possible role for atrazine exposure (Farr *et al*., 2004; Farr *et al*., 2006; Swan *et al*., 2003; Ochoa-Acuna *et al*., 2009; and, Villanueva *et al*., 2005)." The SAP disagrees with this statement. Given the limitations of these studies as discussed in the Panel's responses to Questions 1.1 and 1.2a, the potential for other non-causal explanations is still very high in these studies; they provide only limited evidence of human relevance.

The Panel agreed with the conclusion in the Issue Paper that the experimental toxicology data and not the epidemiologic data are best suited for dose-response and risk assessment at the present time, but that the epidemiologic and toxicologic databases, in combination, can be used to inform hazard identification and characterization. Contrary to the opinion stated in the Issue Paper, the Panel believed that the epidemiologic findings could be used for endpoint selection if the studies are well-designed (even if not achieving unattainable perfection) and experimental data are not available.

Thus, based on an understanding of the mode of action in animals anticipated to pertain to humans, additional research should be focused on selecting appropriate biomarkers of exposure and effect with support from laboratory studies and with special attention to exposure levels of relevance to apical events in human populations. Adequate sample size to achieve sufficient statistical power to detect meaningful effect sizes in the outcomes of interest will be a major consideration in low-dose research, as will adjustment for appropriate known confounding factors for some outcomes, such as maternal age and smoking for low birth weight.

The concerns about effects not yet identified in humans, but seen in laboratory findings are recognized but further efforts in epidemiology will produce additional equivocal findings unless studies are well-designed (i.e., have good exposure and outcome measurement in individual men and women with adequate sample sizes, adjustment for confounding factors, and minimal participation and selection biases) and evaluate health effects of doses that are relevant to real-life human exposures either in the occupational setting (e.g., farm workers) and real-life drinking water exposures (e.g., studying populations likely or known to consume atrazine-contaminated water, perhaps in agricultural settings).

In conclusion, it is not a simple task to develop a satisfactory framework to integrate the results of epidemiologic, clinical and laboratory studies on an issue, such as the health risks of atrazine. At the February 2010 meeting of the SAP, the registrant proposed a framework for a weight-of-evidence approach to the interpretation of findings that bear on human risks of toxic exposures. The proposed framework does not obligate the generation of new data but has the virtue of bringing out assumptions and judgments

that might otherwise be harder to identify. The background materials for this meeting included a paper based on the application of the registrant's proposed framework entitled "Preliminary Review of Recently Identified Atrazine Ecologic and Retrospective Studies" by Breckenridge, Pastoor and Scialli (2010). While the Panel disagreed with a number of the critical judgments presented in Breckenridge *et al*. it would be useful to have a clearer articulation of differences of opinion (and, similarities) between these authors and EPA.

Finally, the Issue Paper is largely correct in its conclusions that 1) the animal data are likely to be relevant to humans, although different mechanisms of toxicity may also play a role in humans and 2) the human epidemiologic data suggest possible adverse reproductive effects. EPA is encouraged to collaborate with other federal agencies or other parties that have funded human studies of these types of outcomes, have biologic samples stored and potentially available for assessment of exposure and have done a good job of investigating the association of the exposure(s) of interest with these outcomes.

Question 2.0: Review of Studies on Mammary Gland Development

At the April SAP, the Agency evaluated effects of atrazine on a number of apical endpoints including neuroendocrine, neurotoxicity, and immunotoxicity effects. At that time, the Agency deferred review of experimental toxicology studies on mammary gland development to the September SAP meeting. Development of mammary tissue occurs in defined stages linked to sexual maturation and reproduction including embryonic, pre- and peripubertal periods, as well as pregnancy and lactation. Given that atrazine affects the hormonal environment and reproductive system, it is not unreasonable to hypothesize that it may also affect mammary gland development. The database of mammary gland development studies includes four studies---three from same the research group (Rayner et al., 2004; Rayner et al., 2005; Enoch et al., 2007) in addition to one conducted by Coder (2010). It is important to note that the key events leading to reported delays in mammary gland development in the young following gestational exposure to atrazine are not known and none of these studies propose a mechanism/mode of action.

Question 2.1

The studies from the two different research groups yielded different results regarding the impact of atrazine on mammary gland development. While the Rayner and Enoch studies report delays in mammary gland development, the Coder study does not. The basis of these different findings is unknown. However, one notable difference in the study design used by the two research groups is the scoring system. The Rayner et al. and Enoch et al. studies used a subjective scoring scale with a 1-4 scoring system (1 = poor development/structure; 4 = normal development/structure) with criteria which vary depending on the age of the pups. In contrast, Coder employed a quantitative morphometric analysis to evaluate mammary gland development. In addition, the Coder study also included the use of BrdU labeling to assess cell proliferation in the mammary tissue.

Please comment on the quality, strengths, and limitations of the mammary gland development studies by Rayner et al. (2004, 2005) and Coder (2010) studies. Please discuss in your comments factors which could lead to different findings. Please comment on the Agency's conclusions regarding these studies on atrazine.

Panel Response

Each of the three studies investigated the effects of atrazine administration to pregnant dams on the body weights, mammary gland development and age of vaginal opening of their female offspring. All three studies used the Long Evans strain of rats and oral gavage as the route of administration. Each study appeared to be well-executed. However, the studies differed in several aspects of experimental design and methodology including: (1) composition of the diet consumed by the dams; (2) light-dark photoperiod employed; (3) inclusion/absence of pair-fed controls; (4) methods used for mammary gland development analysis; and (5) differences in maternal stress levels. Each of these aspects is significant when comparing the findings and conclusions of the studies, which differed in several ways. While each found that atrazine at higher doses (50 mg/kg and 100 mg/kg) affected both dam and offspring body weights as well as mammary gland development in the offspring, the Rayner studies found effects of atrazine on the age at vaginal opening that were not detected in the Coder (2010) study. In addition, the Rayner studies differed from the Coder study in the nature of the effects that atrazine produced on postnatal mammary gland development.

The Panel concluded that numerous significant methodological and procedural differences in the Rayner and Coder studies have confounded the understanding of the toxicological significance of atrazine exposure on certain aspects of female reproductive system development. It was the Panel's opinion that unless/until these differences are resolved, understanding of the effects of atrazine on mammary gland development will be hampered. Consequently, use of the existing data on rat mammary gland development to assess the potential human risk of atrazine is not warranted at the present time.

As noted above, the Panel identified several significant procedural and methodological differences in the Rayner and Coder studies. They are analyzed in detail below:

> (1) Diet. The compositional differences in the diets used in the Rayner and Coder studies may have affected the outcomes of atrazine exposure. The Coder *et al*. study report indicates "The basal diet used in previous studies conducted with atrazine (Rayner, 2004; Rayner, 2005) was PMI Nutrition International, LLC, Formulab Diet® 5008. Compositional differences between the diet used in the previous studies, and that used in the current study, were slight and were not expected to impact animal health or the outcome of the study" (Coder (2010), Page 36 of 2827). However, no data are presented to support this statement. Both Rayner *et al*. studies used Purina 5008 while the Coder study used Purina 5002. These are both natural ingredient diets with relatively high soy content and high and variable phytoestrogen levels. However, they differ in their fat source

(animal fat for 5008, soy oil for 5002), their total protein, fat, and carbohydrate contents, and in some vitamins. There is substantial literature on the effects of soy and phytoestrogens on the mammary gland, as well as on puberty and estrous cycles. While the Organization for Economic Co-operation and Development (OECD) conducted a study of diets of varying phytoestrogen content used in the uterotrophic assay as an endocrine disruptor screening tool to determine their impact on the validation results (Owens *et al*., 2003), it is unclear if this type of comparison has been extended to studies of other endocrine-based endpoints. Based on values reported by others (Brown and Setchell, 2001; Thigpen *et al*., 2003, 2004, 2007) for the isoflavone content of the diets used in the Rayner and Coder studies, the dietary isoflavone levels in the Rayner and Coder studies may have exceeded the levels that Owens *et al*. (2003) concluded could be accommodated without impairing the responsiveness of the test system being validated (i.e., the rat uterotrophic bioassay). There is some disagreement in the literature as to the ability of these diets to cause effects on the timing of puberty, but it may depend on the strain of rodent used. There are sporadic reports in the literature where diet changes have been implicated in altering effects of treatment with endocrine active compounds, and in some cases the diet effect has been attributed to the phytoestrogen content (Boettger-Tong *et al*., 1998; Muhlhauser *et al*., 2009; Thigpen *et al*., 2003; Wang *et al*., 2005). However, there does not appear to be published information on whether the dietary content of soy and/or its component phytoestrogens can shift the dose responses for atrazine. Nonetheless, when looking at the subtle effects on mammary gland development and onset of puberty, these could be important factors to consider.

(2) Light-Dark Cycles. Differences in photoperiods under which the rats in the Rayner (14/10 cycle) and Coder (12/12 cycle) studies were housed have the potential to affect the reproductive system and mammary gland development. Development of reproductive function is known to be influenced by pineal gland activity and there is evidence in the literature that altering light/dark cycles influences the timing of vaginal opening in rats (Lehrer, 1986). There is also literature linking altered photoperiod to elevated breast cancer risk in women (Kakizaki *et al*., 2008) and rodents (Tamarkin *et al*., 1981) and to mammary gland development alterations in prepubertal mice (Mediavilla et al., 1992). However, there does not appear to be published information about whether photoperiod alterations affect atrazine toxicity in mammals. Nevertheless, when considering subtle differences in the effects of atrazine on puberty and mammary gland development in rats, photoperiod differences may be important factors to consider.

(3) Pair-feeding. All of the Rayner and the Coder studies found significant effects of atrazine on dam and pup weights. The Coder study found that atrazine reduced food consumption in the dams in the 100 mg/kg group. They included a pair-feeding control for this dose group which was not included in the Rayner studies. It is increasingly being recognized that maternal nutrition and/or metabolism during gestation can have long-term effects on mammary gland development and

breast cancer risk of offspring. Given that differences in the protein/fat content and composition of maternal diets have been shown to affect offspring mammary gland development (Alvarez-Sanz *et al*., 1986; Eason *et al*., 2004; Hilakivi-Clarke *et al*., 1996, 1998), inclusion of dietary controls is necessary to accurately interpret potential toxic effects of atrazine. Two problems exist in this regard when interpreting the results of the Rayner and Coder studies. First, in combination with reduced food intake at the high dose during gestation, the differences noted in the composition of the diets employed by these two laboratories have the potential of differentially affecting offspring outcomes. Second, conclusions about the effects of atrazine on mammary gland development are complicated by significant differences in the body weights of atrazine-treated dams and their pair-fed controls (Coder (2010), Figure 2), which have the potential of independently altering fetal nutrition and subsequent mammary gland development. Moreover, it is likely the pair-fed controls had major shifts in endocrine and metabolic rhythms in response to the timing of access to their daily food allotment at the beginning of the light-cycle.

(4) Differences in analytical methods. The Rayner studies used a semi-quantitative scoring system that ranked various parameters of mammary gland development including ductal length, ductal branching pattern and complexity, and end-bud number to assess the effects of atrazine. The use of this scoring system has a number of strengths that include: (a) reproducibility, as evidenced by multiple corroborating studies over several years; (b) correlation with known effects of atrazine on vaginal opening that support the efficacy of treatment; and (c) the use of a holistic, or integrative, measure to evaluate mammary gland development. While useful as an approach in a single laboratory with constancy in practice and training, this scoring system lacks the generally accepted standards by which to judge mammary gland development, such as those routinely used by human and animal pathologists. Consequently, inter-laboratory comparisons are difficult and data interpretation is problematic.

The Coder study tried to avoid these problems through the use of quantitative image analysis approaches to evaluate the same development parameters analyzed in the Rayner studies. Although much was made of the differences in these approaches by the Coder study authors, it is surprising that they did not employ both qualitative and quantitative scoring measures of mammary gland development, which would have provided a definitive inter-study comparison. It is plausible that small differences between treatment and control groups existed in the morphometric analyses that, individually, were deemed nonsignificant, but in combination may have impacted scoring, as noted by the Rayner studies. Thus the EPA might consider requesting that the study sponsor have the slides generated in the Coder *et al*. study scored using the Rayner scoring system. Ideally, the same histologists that scored the Rayner slides could score the slides in the Coder study. Direct comparison of results derived from the two methods would be very informative in terms of establishing the biological and toxic effects of atrazine.

(5) Stress. Maternal stress during late gestation has been reported to alter reproductive development in offspring of rodents and primates, including delays in onset of puberty (Harvey and Chevins,1987; Vom Saal *et al*., 1991; Zehr *et al*., 2005). Although maternal stress during gestation has not been linked directly to alterations in mammary gland development in offspring, normal mammary gland development during puberty is dependent upon immune cell function (Lin *et al*., 2002), which is altered in rats born to dams exposed to stress during pregnancy (Klein and Rager, 1995). Evidence was presented in the Coder study that dams exposed to atrazine were aggressive and cannibalized their young. This type of unusual behavior was not observed in the rats in the Rayner studies, suggesting that they were experiencing better control conditions than those in the Coder study. The possibility that stress-related effects account for differences in the results of the Rayner and Coder studies is further supported by the observation that vaginal opening in control offspring of rats in the Coder study was delayed relative to those in the Rayner study, and that atrazine exposure did not produce the expected delay in vaginal opening in the Coder study whereas it did in the Rayner study. Thus, considerations of differences in maternal stress are important when attempting to reconcile the inter-laboratory differences in the biological affects of atrazine.

Question 2.2

In the three studies considered in Question 2.1 (Rayner et al., 2004 and 2005; Coder, 2010), effects have only been observed at an atrazine dose of 100 mg/kg/day. In contrast, Enoch et al. (2007) report delays in mammary gland development as low as 0.09 mg/kg/day following exposure to atrazine and a mixture of several metabolites/degradates (DEA, DIA, DACT, hydroxy-atrazine). The Agency's review of this mixture study is provided in Appendix A (Section A.10) Please comment on the Enoch et al. (2007) study and the degree to which the Agency's review accurately reflects the strengths and limitations of the study. Please include in your comments a consideration of the study design of the mixtures experiment and how this design impacts the interpretation of the results.

Panel Response

The Enoch *et al*. (2007) study was designed to evaluate exposures to environmentally relevant mixtures. The study goal was to determine if a mixture of atrazine parent, its chlorinated metabolites and hydroxyatrazine, in proportions found in the aquatic environment, when administered orally, produces developmental effects in Long Evans rats. Mammary gland development, as assessed by a visual scoring system, was studied in female offspring of dams exposed on gestation days 15-19 to 0.09, 0.87 and 8.73 mg/kg levels of the mixture. Some of the results from their prior studies were used to strengthen the experimental design of this study. Noteworthy is the administration of the mixture at half doses twice daily on GD 15-19. This time period was identified as the critical window in gestation for atrazine's effect on delaying mammary gland development (Rayner *et al*., 2005). The conclusion of the Enoch study

was that acute exposure to mixtures of as low as 0.09 mg/kg causes persistent alterations in mammary gland development of female offspring, and that these effects do not appear to be related to body weight or timing of onset of puberty.

Findings included increased GD20 fetal and PND4 and PND 60 offspring body weights; these may be an example of prenatal metabolic programming (Plagemann, 2005). These observations could be consistent with the Rayner and Stanko (2010) studies at the higher doses if the atrazine parent/metabolites are disturbing a developing brain neurocircuitry. During the public comment period at the meeting Syngenta presented data in adult rats showing that atrazine decreased activation of GnRH neurons in the hypothalamus without affecting paraventricular nucleus (PVN) c-Fos immunoreactivity. Whereas the (presumptively, corticotrophin-releasing factor; CRF) cells of the PVN are integrally involved in the stress response, the distributed CRF system also has CRF cells in the medial preoptic area (mPOA) with synaptic connections to the GnRH cells which could be the vulnerable brain site in atrazine-exposed fetuses (MacLusky *et al*., 1988).

However, there are several limitations in the study design that are relevant to the testing of a mixture. Dose-response data for individual components could have helped clarify the interpretation of the mixture data. The observations were based on administered dose because absorption and bioavailability values were not determined and are unknown. The dose response as a function of 100-fold change in the administered dose of the mixture was weak. No data on chemical analyses of the mixture were provided to determine if the concentration of chemical(s)/metabolite(s) was constant as a function of time (see Panel response to question 3.1).

EPA's concern with the statistical treatment of the data (individual pups versus by dam or litter) was resolved by Dr. Fenton. Similarly, concern for the qualitative scoring of mammary gland development also was ameliorated by Dr. Fenton's detailed explanation of the methodology, noting the long history and widespread use of these measures. It is surprising that the Enoch *et al*. methodology did not include external standards. A head-to-head comparison with multifactor imaging analysis would be a welcome exercise.

The Panel agreed with EPA's conclusion that the Enoch study is not suitable for use in quantitative risk assessment. However, the study does raise some interesting points and dilemmas. Certainly, the very low doses at which effects on mammary gland development were reported could be of importance in assessing the risks of atrazine exposure. However, since these effects occurred at doses so much lower than observed in other studies, there needs to be replication of the study. The observations of this study have not been corroborated using other mixtures of parent atrazine/metabolites reflective of temporal and spatial occurrence or for atrazine parent at equivalent doses. Atrazine itself will, once absorbed, yield a mixture of free and conjugated metabolites (but, little or no hydroxyatrazine) and the proportion of metabolites would likely change temporally. Any replication would need to evaluate the same endpoints measured in this study and, perhaps, other physiological endpoints

Question 2.3

As described in the Introduction of the draft Issue Paper (Section 1.0), the Agency's problem formulation is being conducted in a step-wise manner. One important component of this problem formulation is the evaluation of the overall database and the quality of the available experimental toxicology studies. In the case of mammary gland development, the Agency is proposing to acknowledge these four studies (Rayner et al., 2004; Rayner et al., 2005; Enoch et al., 2007; Coder, 2010) in the hazard characterization but place little emphasis on them in its proposed updates to the dose-response assessment for atrazine. The only findings reported in the Rayner et al., (2004), Rayner et al.(2005, and Coder (2010) studies, occurred at a high dose (100 mg/kg/day) which is approximately 50-fold higher than the PoD of 1.86 for LH attenuation being proposed by the Agency. In the Enoch et al. (2007) study, there are significant limitations with the conduct and reporting the Agency believes which preclude the study from use in quantitative risk assessment.

In light of Panel discussion on Questions 2.1 and 2.2, please comment on the manner in which the Agency has proposed to use the mammary gland development studies in the hazard assessment for atrazine.

Panel Response

The Agency proposed to "place little emphasis" on the mammary gland development studies that were summarized in the Issue Paper, and presented during the meeting. Based on the Panel's own analysis, as detailed in the responses to Questions 2.1 and 2.2, this does seem to be the most appropriate way of handling these studies currently. The changes in mammary gland development were observed only at relatively high doses (100 mg/kg/day), doses to which humans are unlikely to be exposed. No clear dose-response relationship or mode of action has been defined for these mammary gland effects. There are only four papers in the current literature, one of which has significant limitations which make it unsuitable for quantitative risk assessment. Two of the remaining three reports indicate that atrazine delays vaginal opening and mammary gland development until adulthood, based mostly upon subjective endpoints. The third report finds no effect of atrazine on time to vaginal opening or mammary gland development after PND1, and utilizes mostly quantitative endpoints. The results of this latter study, however, are highly variable. Some of the mammary gland responses may be transient; the observation of carryover of effects into the F2 generation needs replication.

The conclusion that the mammary gland effects do not provide a useful PoD for quantitative risk assessment should not be construed to indicate, however, that the mammary gland is not a legitimate target for further analysis. There are good reasons to believe that mammary gland development and mammary gland function are important concerns for environmental agents. These preliminary studies clearly raise the possibility that there may be deleterious effects of atrazine on mammary gland development. Therefore, further study is needed at additional (i.e., lower) doses, with added emphasis

on possible cross-generational effects of parent atrazine and its metabolites, including the hydroxy-metabolites. The mixture study (Enoch *et al*., 2007) is potentially important, but replication of this study and investigation of the individual components of the mixture should be pursued. This mixture study can be explained by any of the following several possibilities:

1) The observed effects are artifacts, owing to a flaw in the study design and/or execution.
2) Only a single very active metabolite is responsible for the observed effects.
3) Multiple very active metabolites may act in a dose-additive fashion.
4) There is significant interaction among the metabolites (i.e., synergy).

The likelihood of significant interaction among these compounds would be slim, but interactions resulting in an order-of-magnitude or greater increases in the toxicity are not outside of the range of possibility. The Panel pointed out that the likelihood of artifact can be evaluated in follow-up studies. If the effects are found not to be artifactual, then additional future studies should address the possibility that one or more very active metabolites exist. Hydroxyatrazine(s) were discussed as candidate active metabolites although they are not significantly produced by mammals and would, therefore, contribute little to the toxicity profile when atrazine is administered in animal models. Overall, these effects are particularly important because they involve developmental exposure and they occur at doses lower than reported for LH surge effects. And, like the LH surge effects, the mammary gland reactions to atrazine are likely to be indirect, involving effects on the hypothalamus or other brain sites.

It remains true that particular features of mammary gland development suggest that it is a valid, and potentially sensitive, target for environmental agents, including potentially toxic compounds such as atrazine. These features include the fact that mammary gland development occurs over several distinct phases of life, including prenatal, pubertal, gestation, lactation and aging; and the breast is a very common cancer target. In addition, the importance of breast milk for healthy development of the young, and the importance of breastfeeding for healthy metabolism of the mother are now well-documented. An additional consideration discussed by the Panel is that the biological mechanisms controlling mammary gland development in the rodent and primates (including the human) map more closely between the species than do those controlling the hypothalamus-pituitary-gonadal (HPG) axis. Concerns raised during the meeting about the applicability of rodent HPG axis findings to humans may, therefore, be less important for mammary gland endpoints.

The comments above argue that developing accepted methodologies and Standard Operating Procedures (SOPs) that would inform useful endpoints related to mammary gland development and physiology will be important. These methods and procedures should address actual lactation endpoints, both qualitatively and quantitatively. This task presents an obvious set of challenges. However, very precise and comprehensive lactation endpoints are routinely measured and used for specific goals in the dairy sciences and industry, where very small differences in lactation outcomes have major

economic consequences. It should be possible to use this rich literature and methodological resource to establish approaches to characterize effects of environmental agents, such as chemical contaminants, on mammary gland outcomes. These methods should, and could, take into account the importance of gathering both animal experimental data and human data that can be meaningfully integrated, and which address cross-generational outcomes. Cancer outcomes were not within the scope of this SAP, but remaining vigilant about breast cancer outcomes is obviously still important.

It is important to consider the weight of evidence in selecting relevant studies as a basis for the evaluation of dose-response relationships for the effect considered critical. Factors to be taken into account include not only the quality of individual studies, including internal consistency of observations, but also the collective weight of evidence. For individual studies, it is always desirable to have several complementary measures of effect, particularly where some are subjective, as a basis for the measurement of internal consistency of the observations. Similarly, consistency of results across studies in the database is an important criterion in assessing weight of evidence for dose-response for critical effects. Effects observed by one research group at doses that are not consistent with the weight of evidence should not unduly influence selection of critical dose-response relationships for the point of departure. This does not imply that the results should not be investigated further, particularly where effects are found at very low doses and where mixtures of compounds that may be synergistic are involved. It is advisable that generic criteria for internal consistency and those relevant to weight of evidence for dose-response for critical effects (consistency, specificity and biological plausibility) are defined *a priori.*

In summary, the Agency's Issue Paper concluded that the few available mammary gland studies do not constitute a sufficient body of evidence to be used in a quantitative risk assessment (i.e., for establishing a PoD). The Panel concurred with this finding. The suggestive evidence in these studies does, however, argue that developing useful and standardized methods for assessing mammary gland functional outcomes remains an important task.

Question 3.0: Proposed Updates to the Dose-Response Assessment

At the April, 2010 SAP, the Panel made two key recommendations related to dose-response assessment: 1) consider pursuing a dose-response analysis based on an internal dose metric, as an alternative to the administered dose metric used in the previous risk assessment; and 2) identify the sensitive endpoint associated with functional impairment and perform benchmark dose (BMD) modeling. The Agency conducted both recommended analyses (Section 5.0 of the draft Issue Paper) in response to these recommendations. LH attenuation, as a key event in atrazine's MoA, is the most sensitive internal dose metric that is associated with a number of endocrinopathies

Question 3.1

Given that there is no robust pharmacokinetic model yet available, a non-compartmental analysis of the existing pharmacokinetic data was conducted and described in Section 5.2 of the draft Issue Paper. EPA thinks this represents the best method to estimate an internal dose metric without exceeding the limits of the available data.

a. Please comment on the Agency's analysis of the pharmacokinetic data contained in Thede (1987) to estimate internal dose reflective of atrazine and its metabolites in rat plasma.

b. Based on Figure 5.5 in the draft Issue Paper, the Agency has preliminarily concluded that plasma levels of atrazine and its metabolites reach (or nearly reach) pharmacokinetic pseudo-steady state in the rat plasma after approximately four daily oral exposures over a wide range of doses. Please comment on the extent to which the data do and do not support this finding.

Panel Response

The Panel's responses to Question 3.1a and 3.1b have been combined.

The key uncertainty in the characterization of exposure for dose-response analysis is the extrapolation to low dose in humans from high dose studies in animals. The "best available science" for doing these extrapolations is physiologically-based pharmacokinetic (PBPK) modeling, particularly when a specific target organ has been identified (e.g., the hypothalamus-pituitary complex, for which there are adequate data to identify the relevant dose metric associated with the critical effect). PBPK models are clearly not the quickest or easiest tool to apply. While the Panel was cognizant of the time and resource constraints on the Agency with regard to conducting risk assessments of various pesticides for re-registration or other risk management activities, the Panel found it disappointing that PBPK model refinement for atrazine has not been undertaken since 2007 when the McMullin model was published in the peer-reviewed literature. PBPK models have been or are being developed for other pesticides (e.g., synthetic pyrethroids, chlorpyrifos), and for atrazine by the registrant. It would be prudent for the Agency to work with the current model described by the registrant and make appropriate modifications. It is recommended that the Agency be more proactive in this area, and that it begin to address more rigorously the uncertainties of extrapolation from rodent studies to low dose regimens in humans.

In the absence of a functional PBPK model and appropriate data, the Panel very much supported the work of the Agency in pursuing a dose-response analysis based on an internal dose metric, as an alternative to the administered dose metric used in the 2003/2006 risk assessment. This should reduce the uncertainty associated with inter-route, interspecies and intraspecies (inter-strain) extrapolations. In the non-compartmental analysis, the Agency has attempted to make maximal use of the limited

data that are available. EPA has made mainly conservative choices in the absence of empirical information and verified estimates to the extent that is possible.

The Panel agreed with the Agency that, on the basis of the currently available data, plasma appears to be a reasonable biological compartment that is reflective of tissue dose, and that use of area under the plasma concentration time curve (AUC) provides an appropriate measure of internal exposure. It is the best way of approaching the problem of relating the exposure of rats to atrazine to the magnitude of resulting biological responses, and extrapolating to the prediction of human exposure levels. This is particularly useful in the absence of knowledge of the exact site of action since it provides a measure of the opportunity for distribution to all tissues (target and non-target). This conclusion is supported by the work of Stoker and Cooper (2007). The studies of Dooley and co-workers demonstrating the covalent binding of DACT to proteins with available cysteine residues in all parts of body (19 in rat pituitary (Dooley, *et al*., 2008)) 30 in rat brain (preoptic area, medial basal hypothalamus, and cortex; Dooley, *et al*., 2010) indicate the potential impact of DACT on a range of proteins (enzymes and structural components). However, there is currently no evidence to identify the consequences (inhibition or activation) of binding on the activity of individual proteins.

In the absence of good data on the plasma concentrations of parent compound and individual metabolites at various elapsed times after dosing, the use of total chlorotriazine based on total ^{14}C-compounds is a reasonable first step (particularly in absence of information on the pharmacodynamic activity of the parent compound and individual metabolites). However, the data of Thede (1987) has some limitations, and these should be kept in mind when interpreting any pharmacokinetic analysis. The plasma measurements are based on total radiolabel, and the samples were taken at 24 hours after each of the repeated doses. This means that the plasma profile is poorly defined, and, because of the widely separated sample points, will appear smooth. There are too few measurement points in the washout period to obtain sound estimates of elimination parameters. The Agency's interpretation of the Thede data includes the assumption that the parent compound and individual metabolites (DACT, DEA, and DIA) are all equally active toxicologically. Considering that atrazine is short-lived and that the glutathione conjugates are likely to be inactive, several possible dose metrics are available. These include sum of total metabolism of atrazine, total metabolism minus glutathione conjugates, total of DIA, DEA and DACT, total of DIA and DEA.

Atrazine and its two mono-dealkylated analogues are rapidly depleted by primary detoxication; the major end product (DACT) dominates the plasma profile. There is an assumption in the interpretation that none of the pharmacokinetic processes (absorption, distribution, metabolism and elimination) is saturated, and so the proportions of the parent compound and metabolites are independent of dose. The composition of the total radiolabel will change with time after each administration by oral gavage. Immediately after dosing, the concentrations of atrazine, DEA and DIA will be high, and will fall over the following hours, and after 24 hours (the time of the next dose) more than 90% of the AUC will again be made up of DACT. In this context, it is recommended that the

differences in molecular weight between atrazine and its metabolites (215.7, 145.5, 173.7, and 187.7 for atrazine, DACT, DIA, and DEA, respectively) are accounted for in the calculation of milligram equivalents of atrazine. The factors involved (1, 1.48, 1.24, 1.15) will be important, especially at later times. Evidence in Laws *et al*. (2003), Stoker *et al*. (2002) and McMullin *et al*. (2004) might be used to clarify the Agency's approach of assuming equal toxicological potency for atrazine and all of its metabolites.

Estimates of the elimination rate constants on the basis of the last three points (based on a single animal) are associated with some uncertainty. However, there is a very rapid fall from the level at the time of the last dose. Indeed, if more frequently-collected samples were available, the concentration would have been seen to rise after the last dose before falling from the peak level. This pattern is indicative of two simultaneous processes, one very rapid that is virtually complete over 24 hours, and then a slower process. On the other hand, the data from the new registrant study (summarized in Breckenridge, *et al*., 2010), where there are more points available over the washout period, do not indicate double exponential elimination. It would be worthwhile to look closely at the washout data, since the estimates of the elimination rate constants will be important in any predictions or extrapolations to other dosing regimens.

If AUC of plasma concentration is to be used as the measure of internal dose of triazine equivalents, then consideration should be given to the interspecies differences in bioavailability, rate of glutathione conjugation, role of intestinal metabolism, and red blood cell/plasma partition coefficient. Furthermore, knowledge of the contribution of elimination processes to the internal exposure profiles is essential, and work is necessary to assess the uncertainties associated with this when extrapolating between species (e.g., rat to human). The difficulties involved are highlighted by known differences between strains within a species, as exemplified by the differences in the LH response, and biotransformation enzymes between Long Evans and Sprague Dawley strains of rats. Long Evans rats have lower overall CYP activities, which may lead to slower clearance (Jori *et al*., 1971). The Panel recommended that studies to provide parameterization for a Long Evans PBPK be initiated as soon as possible. This parameterization should include an intestinal biotransformation metric (either microsomal or Caco-2 cells). The hepatocyte data provided by the registrant were useful, particularly given the differences between rat and human metabolite profiles, and differences between microsomal (Joo *et al*., 2010) and whole hepatocyte results in the formation of DACT (the major metabolite produced *in vivo*). *In vitro* studies also may be conducted with mixtures of the metabolites in a range of concentrations that bracket those in the environment. Such data could be used in the next risk assessment.

It was the opinion of the Panel that the approach taken by the Agency is the correct one, but it requires better data than those currently available. The pharmacokinetic study in progress as described in the registrant's presentation at the meeting and in Breckenridge *et al*. (2010) will provide data of the required quality since the use of frequent sampling and mass spectrometric detection of the analytes is providing tight definitions of the PK profiles of the individual compounds (parent compound and analytes). This type of study lends itself to analysis of relationships between the AUCs

of the different compounds and biological responses. This information combined with good estimates of plasma partition coefficients should allow the development of PBPK models that can be used to extrapolate from rodent to other species, including humans. This would provide a sound basis for the setting of benchmark doses, and PoDs, providing that it is developed for the rat strain in which the PoD or dose-response is characterized. While it is clearly preferable to have a validated PBPK model, within the time constraints for regulatory programs, it may be necessary, at a minimum, to prioritize collection of data that would reduce some of the uncertainty associated with the current Agency analyses. Certainly, the outstanding issue is how to deal with the uncertainties associated with the constraints of the existing data. Despite the limitations of the existing data, the exercise has been extremely valuable in informing critical aspects relevant to the risk assessment and monitoring programs.

Pseudo steady state implies that the absorption and distribution processes are matched by the elimination processes. The interpretation of the levels presented in Figure 5.5 of the Issue Paper is misleading since all plasma samples were taken at 24 h after the previous dose. The points defining the profiles in Figure 5.5 are the minima in the saw tooth profile, and not the average levels. Perhaps if the samples had been taken at 12 h after dosing, the profile would have been different. If the samples had been taken at random times after dosing, there would not have been such a smooth plateau. The real profile will not have such small amplitude fluctuations as illustrated in Figure 5.4 of the Issue Paper (see Figure 6 in Breckenridge *et al*. (2010)). This means that the AUC is underestimated. The bias will be approximately constant across the plateau area, and so it will not change the goodness of fit of AUC versus administered dose, and could give a false sense of security.

In the PK experiments undertaken by Thede (1987), the same dose was used throughout, and under these conditions, the approach taken of estimating AUC provides a useful measure of exposure. The estimate of AUC in the region of pseudo-steady state exposure over the full 4 days is proportional to oral gavage dose, as shown in Figure 5.6 of the Issue Paper. However, as described above, caution needs to be taken in the interpretation of the Thede (1987) data set because of the poor definition of the serum profile. The new registrant PK data sets show that there are large fluctuations in serum concentrations of atrazine and its metabolites with time after oral gavage dosing, and that a smoother profile is achieved with dietary administration. Despite the difficulties in estimating dose consumed in the diet compared with administration by oral gavage, it is worthwhile. Dietary administration provides a smoother plateau, is less stressful and disruptive than gavage, and is more realistic representation of human exposure through drinking water, depending on the presence of food in the digestive tract.

c. Figure 5.8 of the draft Issue Paper shows a plot of LH attenuation data versus administered atrazine across a range of exposure durations (i.e., four days up to six months). The data contained on this figure were collected from several different laboratories and involved two modes of oral administration (dietary, oral gavage) and two different rat strains. Attenuation of LH, as measured by percent of control, is remarkably similar across studies, strains, laboratories, and most notably duration of

exposure. The Agency has preliminarily concluded that the findings on Figure 5.8 strongly support the hypothesis that pseudo-steady state is achieved (or nearly achieved) in adult rats after four daily consecutive oral exposures. The Agency also believes that pseudo-steady state is strongly associated with the attenuation of the LH surge following atrazine exposure in rats. Please comment on the analysis shown in Figure 5.8 and the Agency's preliminary conclusions related to these findings in rats.

Panel Response

The Panel consensus was that Figure 5.8 is an imaginative distillation of data from two different rat strains across 15 years of data collection from four laboratories. The figure serves as a visual meta-analysis and assists in the understanding of the atrazine dose-related attenuation of the LH surge. It is remarkable that the overall response is similar given that there was such diversity in the study paradigms: different rat strains and different dosing regimens. The model presented in Figure 5.8 represents a good, initial approach to data integration and has great utility for exploring concepts and formulating approaches to test hypotheses further.

Overall, the Panel agreed that several aspects of the Figure need refinement in terms of the axis scaling, winnowing of LH data, and inclusion of the higher doses to incorporate the full range of data. All these aspects can be corrected and may yield a more accurate and sensitive model.

The first concern is the selection of a linear scale for the X and Y axis. In the Figure, the response data have been plotted as percentage of control against the linear dose (mg kg^{-1} d^{-1}) of atrazine. A clearer and more standard approach would be to plot the response data (including confidence intervals) on a standard normal probability scale (i.e. in standard deviation units) and the concentrations on a log scale. This would then facilitate including the higher doses in some of the studies summarized here as well as other studies which included single doses (see Figure 5.2 from Cooper *et al*. (2000)). When interpreting this plot, it needs to be borne in mind that the points near the limits (0% and 100%) are associated with larger variance (and greater uncertainty) than points near the median. It would be useful to incorporate the higher doses to see where they fall. This should not be precluded on pharmacokinetic grounds, given that AUC is highly correlated with administered dose up to a concentration of 100 mg kg^{-1} d^{-1}, with no indication of saturation of any of the pharmacokinetic processes (absorption, distribution, metabolism, and elimination). Further, the higher doses fitted on the Hill model (Figure 5.11 of Issue Paper). If there is no effect of further increasing the dose, then this would imply that a limiting response has been reached; this would be of interest in setting bounds to the dose response space. Alternatively, if there is an effect of further increasing the dose, it is possible that two phases of PK dynamics can be characterized. Inclusion of the complete data set will help to further develop the hypothesis. This may allow for linking higher doses with more dramatic attenuation of the LH surge and more immediate effects of downstream fertility outcomes.

The second concern is the selection of 1800 h LH values as the basis for calculating "Percentage of control for the LH Surge Attenuation" for the Y axis rather than peak LH or AUC levels. Inspection of the raw data of Morseth and Minnema revealed that peak LH varied between 1600 – 2000 h in individual animals and treatment groups. For example, in the Morseth 1-month study (1996a), for both decapitation and repeated sampling studies, the peak of mean LH concentrations in the controls (0 mg/kg/day) was at 1800 h, but in the 5 mg dose groups, the mean LH peak was at 1600 h in the decapitation study and 2000 h in the repeated sampling study. In the Morseth 6 month study (1996b), LH appears to peak at 2000 h in controls and at 1800 h in the 25 ppm group in repeatedly sampled rats, whereas in the decapitation study, both 0 and 25 ppm groups peak at 1800 h. When assessing the LH peak, the best LH values to use, if serial samples are available, are the actual LH maxima, regardless of the time at which they occurred.

Third, of even greater concern is the variability of the LH data. Data from the Cooper (2010) 4-day study provide the major proportion of the dose response data. In this study, groups of Long Evans rats were sacrificed at various times in proestrus. Verification that the rats were in proestrus was based upon three criteria, namely, a proestrous vaginal smear, ballooned uterus, and an elevated serum progesterone concentration determined *a posteriori.* Only data from those rats that met all three criteria were used in the analysis. Additional data for dose levels between 1.5 and 5.0 mg/kg/day were obtained from the Morseth (1996) and Minnema (2001) studies that used Sprague-Dawley rats. These data are much more variable than the Cooper data (see below); therefore, there are no apparent differences among the means, or among the NOAELs or LOAELs.

There are significant problems with the manner in which the LH data in the Minnema and Morseth studies were analyzed that can be corrected. Both of these reports employed all LH values from all samples that were assayed. However, precise determination of an LH surge (elevation of LH above the 95% confidence limits of baseline for at least 2 consecutive time points) was not exacted. Not all animals had an LH surge (including controls); therefore, there is a significant underestimation of LH values, means and overestimation of standard deviations. It was not possible to determine whether LH surges were defined in the McMullin report.

Correction of the Minnema and Morseth data sets such that only those animals that exhibited an LH surge are included in the analysis of LH surge peaks, will increase the mean LH levels, decrease the variation in the data, and might also alter the conclusions with regard to the NOAEL dose. Most importantly, correcting the data will affect the benchmark response (BMR) analyses.

A fourth concern is that the potential importance of the correspondence between the dose of atrazine and the proportion of animals with delayed/ lengthened cycles was overlooked. The delayed cycles are probably the result of insufficient or absent LH surges. For example, the percentage of atrazine-treated animals that did not meet the above criteria for an LH surge for each dose in the Minnema paper were: 0 mg, 34.2%;

2.5 mg, 30%; 5 mg, 50%; 40 mg, 67%; 200 mg, 58%. These proportions are high, i.e., about one-third each of the vehicle and 2.5 mg groups did not surge on the expected day of proestrus. More importantly, there was a trend towards increasing proportions of animals that did not surge with increasing doses of atrazine. Similar results were not reported by Morseth, but the data are available. This endpoint should be reported and analyzed separately for all studies. In this regard, a similar relationship was observed in the new Cooper study (2010) and registrant data, both of which indicated that the proportion of animals that failed to meet the proestrous criteria or have lengthened cycles, and therefore failed to surge, increased with increasing atrazine dose, although the proportions are much lower than in the Minnema data. This is a very important functional observation, especially as attenuation of the LH surge has no adverse effect on reproductive function and does not prevent ovulation until about 80% attenuation (Krieg, *et al*., 2000; Kirchick, *et al*., 1978). Therefore, the proportion of animals and the latency to exhibition of delayed cycles might constitute a better endpoint, or "adverse response" for determining the effect of atrazine than is attenuation of the LH surge.

Finally, data from two strains of rats with significantly different sensitivities to atrazine are combined, but they should be also be evaluated separately. Moreover, it is important to recognize that the appearance of the curve may change with data obtained from exposures at different stages of the life cycle (i.e., pre-pubertal/peripubertal and aged/peri-senescence). As additional data, particularly with high(er) atrazine doses, are added to the model, a mindful consideration of non-LH effects should still be considered. For example, food intake and body weight changes (as % control), described in the Morseth and Minnema studies, plummet at the high atrazine doses and generate a parallel, inverse curve.

It is also important to consider the pattern and duration of exposure to atrazine when considering the data from the various studies summarized in Figure 5.8. The Panel questioned the Agency's conclusion that pseudo-steady state is strongly associated with the attenuation of the LH surge following atrazine exposure in rats. Although the PK data indicate that a pseudo-steady state has been achieved over the four days of dosing in the Thede experiments, it is important that the idea of a pseudo-steady state being necessary for attenuation of the LH surge is not used in interpreting these data. The data do not support this interpretation since no other dosing pattern was used. In order to be sure that a pseudo steady state was necessary to produce an attenuation of the LH surge, it would be necessary to examine other dosing regimens (for instance a high dose on day 1, a low dose of day two, no dose on day 3 and a high dose on day 4) which gave the same overall AUC of atrazine and metabolites as in the reported experiments. A series of such experiments in which the biological response corresponding to the different patterns of dosing was measured would provide evidence for or against the assumption that pseudo steady state was necessary to produce attenuation of the LH surge at low doses. Until such evidence is available, it is necessary to keep an open mind since it may not matter whether the system is in steady state or is subjected to concentrations fluctuating from day to day. There may be a critical window of exposure that is less than the four days used in the reported work, and providing that a minimum exposure (AUC) is maintained over this period, a reduction in the LH surge will result. It has been assumed

that the tolerance remains constant throughout the four days of dosing. If a range of higher doses (corresponding to higher AUC) is used, it may be possible to see whether there is an upper limit in exposure beyond which no increase in suppression of the LH surge occurs.

Question 3.2

In Section 5.3, the Agency describes a benchmark (BMD) analysis conducted for LH attenuation studies. As described in the draft Issue Paper, the Agency initially considered a wide range of toxicological endpoints for inclusion in the BMD analysis. In considering all of the data, this analysis focused on studies measuring LH attenuation where female rats were exposed orally for four days and/or longer. This was not only the most sensitive endpoint but its temporal and dose response are well defined experimentally. The LH effect originates from atrazine's effect on the hypothalamic control of pituitary function through its interference with GnRH neurotransmitters. Thus, protection from LH attenuation as a precursor event to several functional impairments would be protective of atrazine's neuroendocrine effects.

a. Please comment on the BMD analysis summarized in Section 5.3 of the draft Issue Paper with details provided in Appendix C.

Panel Response

Mode of Action; Biological Significance of the Selected Critical Effect/Degree of Protection - The Agency is to be commended for considering early key (i.e., precursor) events in hypothesized modes of action as the basis for characterization of the critical dose-response relationships as more relevant than the apical endpoint to prevent adverse effects in the human population. The data on these earlier precursor or key events (attenuation of the LH surge, in the case of atrazine) often lend themselves to better characterization of dose response owing to the availability of incidence data at more dose levels than have been traditionally included in standard longer term toxicity studies. The hypothesized MOA for atrazine's effects on reproductive function is presumed to include suppression of GnRH pulses, consistent with an immediate downstream effect on the LH surge.

Based on the available data, a benchmark dose modeled from data on suppression of the LH surge appears to be protective for other endpoints, since this phenomenon occurs at doses lower than for the wide range of effects identified in a rather extensive toxicological database. The Panel agreed with the Agency's conclusion that, at the present time, the most sensitive biological endpoint is attenuation of the LH surge.

It must be noted, though, that while attenuation of the LH surge over one estrous cycle represents a sensitive biological response, a causal association between the observed modest, yet statistically significant, changes in LH surge peak and adverse fertility measures has not been made. The relatively small attenuation of the LH surge

may well be a harbinger of a possible future adverse event, but at present this remains to be demonstrated and other possible endpoints should continue to be investigated.

Specifically, it should be noted that only about 20% of the LH that is released on proestrus is required for normal ovulation and fertility; therefore attenuation of the LH surge to about 40% of control levels, observed at doses up to 30 mg/kg bw/day (Figure 5.8 of the Issue Paper), would not be expected to lead to subsequent loss of reproductive function. Indeed, there was no effect on estrous cyclicity in atrazine-treated groups at doses of 30 mg/kg bw/day or less.

The possibility exists that attenuation of the LH surge at a level less than that which prevents ovulation might not be protective for other reproductive effects, such as anovulation or decreased fertility. There are also some data seemingly inconsistent with the hypothesized mode of action. For example, suppression of GnRH pulses would be expected to lead to an increase in serum follicle-stimulating hormone (FSH) levels. There is no evidence to show that this is the case, even at high doses of atrazine (Cooper *et al*., 2000; Foradori *et al*., 2009). Indeed, it appears to have been assumed that the effects of chronic exposure to doses less than those that would alter reproductive function may have adverse effects on other areas in the brain and be potentially deleterious to other functions. However, these sites of action and responses other than LH surge attenuation remain to be identified, perhaps using molecular approaches. For example, binding to the growth hormone releasing hormone (GHRH) receptor and decreases in GH release following exposure to atrazine has been suggested based on a recent report by Fakhouri *et al*. (2010). In relation to relevance to humans, changes in basal LH release, which also are directly regulated by GnRH secretion, might also be a sensitive endpoint.

The biological significance and degree of adversity of the effect used to derive the benchmark value remains to be addressed robustly and needs to be considered carefully in relation to other endpoints on continued exposure, based on the totality of data, including that from more traditional studies. This would permit consideration of the degree of protection offered by the proposed point of departure compared with those based on traditional apical toxicological endpoints, for which the database is considerable. This will need to be taken into account when, for example, selecting the appropriate uncertainty factors, along with the adequacy of the toxicokinetic and toxicodynamic data that will inform the inter- and intra-species extrapolations.

It may be helpful, then, to derive and array a variety of potential points of departure for relevant endpoints with a view to characterizing their degree of protection (in relation to biological significance of the observed effect) and associated uncertainty. This would serve as a critical step for the interpretation of the derived BMD, not just to provide understanding of the biological significance of observed effect, but also as a tool to compare and contrast the uncertainties associated with various options. Comparative uncertainty analysis would include consideration of appropriate interspecies and intraspecies adjustments, in the context of their associated degree of uncertainty.

Adequacy of Data for the Benchmark Dose (BMD) Analysis - The advantages of deriving and using benchmark doses rather than the historical effect/no-effect levels are well known and documented; BMDs reflect the use of all of the dose-response data in a much more robust fashion, accounting for dose selection, dose spacing and sample size. They also offer potential advantage from the perspective of integration of data from multiple studies.

The existing data on the dose response of the attenuation of the LH surge associated with exposure to atrazine (in rats) are sufficient to permit robust analysis for the benchmark dose. These include significant data in several strains exposed to a range of doses delivered either by oral gavage or the diet.

Selection of the Critical Study for BMD Modeling - The Agency modeled four relevant candidate data sets, representing the four studies summarized in Figure 5.8 and, appropriately, selected the "new" (unpublished) NHEERL 4-day data as the most appropriate for deriving a PoD (i.e., Cooper, 2010 (unpublished).

Selection of the Cooper (2010) study as the basis for benchmark dose modeling is appropriate, not only because of the range of doses used, and the lower variability of the data, but also owing to the inclusion of data on animals experiencing a statistically defined LH surge. This was not the case for the Minnema and Morseth studies. For these studies, if one were to reassess the data based only on animals having a statistically defined LH surge, the overall variation in the results would decrease and the average LH values would increase. For example, in the Minnema data set, the LH levels at 1800 h in the control, 2.5 and 5.0 mg/kg/day dose groups are reported as 2.73, 2.95 and 3.38 ng/ml; if corrected, they would be 4.08, 3.78 and 4.43 ng/ml, respectively, with less variation about the means. The latter values were calculated using data only from animals whose serum LH levels increased greater than 2 standard deviations (SD) above baseline levels for at least two consecutive times, a commonly used definition of an LH surge.

Prior to the meeting, in public comments submitted by the registrant, certain aspects of the new Cooper *et al*. (2010) data were subjected to criticism which, if accurate, could have negated the value of the study. The Panel had the opportunity to hear Dr. Cooper's response to these criticisms. One of the criticisms was that the statistical analysis did not include the control data from Study Block III. Dr. Cooper clarified that this assertion was untrue, and that this data set was, in fact, included in the analysis. A second criticism was that the age range of the animals studied was too great. Dr. Cooper clarified this issue when he explained that the age range covered was valid in that each animal fitted the selection criterion of maintaining an estrous cycle over the course of the study period. The third criticism was that each of the three study blocks was separated by too much time. Because of this, the three data sets should not be analyzed together. The Agency BMD analysis incorporated the data from all three study blocks. Since a purpose of these three studies, taken together, was to identify (initially) effect levels, and then, ultimately, no-effect levels (which were identified in Study Block III), an approach that could be taken to address the third point would be to attempt a BMD analysis using only the data from Block III and compare the outcome of this analysis with

that which considered the data from all three study blocks combined. The Panel concluded that these new data were appropriate for use in benchmark dose analysis.

Other Aspects - The only data that supposedly were used in the modeling of the dose response curve for attenuation of the LH surge by atrazine were from the new Cooper *et al*. study. However, in Figures 5.9 - 11 of the Issue Paper, data from other studies were presented, but their source is unclear. In Figure 5.9, the highest dose is graphed at 75, but the Cooper *et al*. highest dose was 25. The BMD is about 5. In Figure 5.10, the lowest dose is <1, a dose not used in the Cooper *et al*. study and the highest dose is about 13, which also was not a dose used by Cooper, *et al*.. The BMD level in this figure is about 2.1. In Figure 5.11, the dose range extends to >300, the lowest dose only around 10, leading to a BMD of 50. Documentation for these additional data sources should be included in the Figure legends and accompanying description.

Traditionally, the 95% confidence limits, which is 2 SD from the mean, are used to determine whether there is a difference in LH levels from the controls. In this case, the BMDL was selected for use as the PoD; it was defined as the 95% lower confidence limit, which appears to be in agreement with traditional approaches.

In Figures 5.9 - 11, it is not totally clear whether the BMDL, which is the 95% lower confidence limit, is, in fact, any different from the BMD, which is defined as 1 SD (which would be the 67% limit) from the mean of the controls. In each figure, the BMD and BMDL are equidistant from the mean.

b. The selection of a suitable benchmark response (BMR) is an important component of conducting a BMD analysis. As described in Section 5.3, for continuous endpoints, like LH attenuation, the BMR most often represents an X% change from background levels (or untreated controls). Typically, the BMR is selected on the basis of a combination of biological (mode of action, quantitative link between key events and adverse outcomes, historical/concurrent controls) and statistical considerations (sample size, variability, etc). LH attenuation may be potentially associated with several adverse outcomes and the level of attenuation of the LH surge considered to be adverse is a function of several factors including the functional outcome and the life-stage. There is an absence of information concerning the level of response (or % change) associated with the various functional impairments. Also, the differences in reproductive cycles/aging between rodents and humans add an additional level of complexity to establishing a specific BMR value. When an X% change can not be defined, the Agency's draft BMD guidance suggests that the BMD and BMDL corresponding to a change in the mean response equal to one standard deviation from the control mean be used as the BMR. This approach was applied to atrazine. Please comment on the scientific factors important for establishing a BMR for LH attenuation as part of the BMD analysis for atrazine.

Panel Response

A solid understanding of a benchmark response is needed for BMD analysis. In this case, attenuation of the LH surge is a sensitive measure of atrazine exposure. It

directly relates to mode of action, being one of the key precursor events to the ultimate endpoint(s) of toxicity and it exhibits dose response characteristics. Thus, it is a suitable benchmark response.

Even so, inherent difficulties exist. These include the wide range in LH surge amplitude normally observed in both rodents and women, and the variability in the exact time of day the actual peak in LH surge occurs. It should be cautioned that, when determining the BMR, inclusion of highly variable data, such as averaging data from animals exhibiting a surge with those not exhibiting a surge, could lead to an overestimate of the BMD. With regard to demonstration of a dose response of inhibition of the LH surge to atrazine exposure, several replicates of the dose response curve, ideally in each strain of rat tested, would help to validate the precision of the BMR determination. It has been well documented that the Long Evans strain is more sensitive than the Sprague Dawley strain, and therefore data from these two strains should not be combined for the determination of a BMR. Their differences should be assessed, acknowledged and taken into consideration, but their data should be analyzed separately.

There are ample data in the literature to demonstrate that complete (or nearly complete, i.e. ~ 80%+) inhibition of the LH surge does directly correlate with adverse fertility outcomes, including altered cyclicity and inhibition of ovulation. Several studies presented in the Issue Paper, using both oral and dietary exposure to atrazine, have demonstrated very significant reductions of the LH surge and resultant effects on fertility measures, such as cyclicity, tightly linked in time. These tightly linked changes in LH endpoint measurements (decreased LH surge amplitude or AUC) and associated modification of fertility measures were observed with high doses of atrazine.

Outside of these very dramatic changes in LH surge (at or near complete abrogation), data correlating more modest, yet statistically significant, changes in LH surge characteristics with key fertility events are lacking. These data are not only lacking from the series of studies presented in the Issue Paper, but importantly, there are not sufficient data in the broad body of literature (rodent or human) related to reproduction and fertility to directly link small or moderate changes in LH surge characteristics with fertility outcomes. This does not negate the possibility that they do exist. Furthermore, a moderate change in LH surge peak amplitude in the context of one cycle may impart minimal to no effect on overall lifetime fertility measures. However, it is important to note, modest changes in LH surge amplitude over long durations of the reproductive lifespan are likely to have an effect on overall fertility measures. This potential should be addressed experimentally.

Specific data which examine long-term diminution of the LH surge over the reproductive lifespan and resultant effects on fertility have not been studied with atrazine. In fact, there is little or no understanding of the long term effects of LH surge diminution and fertility outcomes following exposure to any compound or agent known to modulate the LH surge. In this regard, the responsiveness of the hypothalamic-pituitary-gonadal axis to increasing estradiol and the ability of increasing estradiol to trigger the ovulatory surge of LH are reduced as the aging process progresses in both rodents and women. In

this context, changes in the characteristics of the LH surge including decreased amplitude of the surge in both rodents and humans precede cessation of cyclicity. Thus, modest changes in LH surge amplitude elicited by atrazine exposure together with the normal biological changes occurring during the aging process may result in overall enhanced sensitivity to the effects of atrazine. Thus, atrazine exposure during the reproductive aging phase of the life cycle may elicit greater effects on the LH surge and fertility. Studies specifically addressing the effects of atrazine during all key phases of the reproductive lifespan are lacking. Additional uncertainty arises with the significance of the duration of exposure.

Data are clear in identifying that a greater-than-one pulse of exposure to atrazine is necessary for attenuation of the LH surge. For example, single high doses (over 100 mg/kg) administered on the morning of proestrus did not alter characteristics of the LH surge occurring later the same day. Additional data clearly demonstrate a once daily dose for 4 days and beginning on estrus can induce significant inhibition of the LH surge peak. In this instance, a dose response is observed. However, what is not clear is if less than 4, but greater than 1 days' exposure is sufficient to alter the LH surge. Further complicating the matter, it is not clear if a 4-day exposure, beginning on a different day of the cycle, will result in changes in the LH surge similar to those when dosing begins on the morning of proestrus. Understanding of the relationship between duration of exposure and phase of the cycle will be key in translating rodent data to humans for risk assessment purposes.

It should be understood further that the significance of the duration of exposure and phase of the cycle may be difficult to define in the rodent system. Although there are similarities between elements of the estrous cycle in a rodent and the menstrual cycle in women, differences do exist also. The overall duration of the estrous cycle (4 days) may be too short to define specific cycle-phase related sensitivities.

Thus, based on the Panel's current understanding of effects related to atrazine exposure together with its level of understanding of long-term effects of modest changes in LH surge and fertility, they concluded that the available data are insufficient to directly relate a specific percent change in either amplitude or AUC of the LH surge with fertility outcomes.

The alternative, as presented by the EPA, to use the mean and standard deviation approach is basically a statistical means to define a significant change. This approach represents general, default guidance. It must be remembered that use of this approach inherently has the same limitations related to the inability to identify a specific percent change in the LH surge as a key benchmark response. Furthermore, there is some question if the use of one or two SD would be appropriate. The Agency could examine the variation in historical LH surge data to determine the appropriate SD to define change. Overall, use of the mean and SD approach is a sound and logical approach by the Agency at this time and will allow the Agency to meet current time constraints.

However, given the protective capacity of the response, the uncertainties associated with LH signaling differences between rats and humans, as well as the

uncertainties associated with "how much diminishment" is necessary for the occurrence of an adverse reproductive or developmental effect, calibration of this response with "down-stream" apical reproductive/developmental endpoints should be carried out to provide a linkage between this "early event" and the adverse effect. In this way, a putative numerical reduction in LH surge (i.e. ~30%) could be related to an apical endpoint of reproduction or development. These calibration studies should be conducted with an appropriate positive control (i.e., a known GnRH antagonist). Additionally, route of exposure should be by diet to more closely mimic the most relevant pattern of exposure in humans.

At the present time, use of data on attenuation of the LH surge is a suitable benchmark and directly relates to the present understanding of mode of action. However, additional endpoints including but not limited to growth hormone activity should not be overlooked. Furthermore, consideration of additional endpoints and/or additional modes of action/key events may allow for integration of both cancer and non-cancer health effects of atrazine.

Question 4.0: Approaches To Evaluating Water Sampling Strategies And Frequency Of Monitoring

The ultimate utility of pesticide monitoring data for drinking water exposure estimations in human health risk assessments depends on how well the monitoring data characterize the spatial and temporal variability of pesticide concentrations in drinking water, with an emphasis on the upper end of the exposure distribution. In addition, the method used to estimate concentrations from monitoring data depends on the duration of concern and how critical it is to estimate peak concentrations.

Question 4.1

A well-designed drinking water monitoring study takes into account both spatial and temporal patterns of exposure. Please comment on the USEPA's recommended framework for designing a monitoring study by targeting the most vulnerable areas, targeting seasonal times for more intensive sampling, basing sampling frequency on the toxicological exposure duration of concern, and using autosamplers to supplement monitoring data (see Section 7.2 of the draft Issue Paper).

Panel Response

The Panel noted that if the true purpose of a monitoring program is to generate valid "exposure estimates," then sampling the finished water of a community water system (CWS) is more appropriate than sampling its raw input water. Studies that have examined the transport of atrazine through a water treatment plant sometimes show a statistically significant reduction in atrazine between the source water and finished water; however, complete removal is rare and some water treatment plants show no reduction (Coupe and Blomquist, 2004; Kingsbury *et al*., 2008). This suggests that the material covered in Section 7.3.2 of the Issue Paper might be better used in Section 7.2 where the

"General strategy for designing a monitoring study to characterize DW exposure" is discussed. However, if understanding the fate and transport of atrazine *to* a CWS is desirable, then sampling the raw water source is more appropriate.

There is a clear need to group drinking water facilities into categories according to how atrazine is delivered to the CWS. The distribution of atrazine in a water body changes significantly depending upon from where the CWS draws its raw water, whether delivery to the system incorporates some form of holding pond or reservoir or is directly from a stream or river, and whether mixing of raw water sources occurs. A one-size-fits-all monitoring network does not seem appropriate to address the estimation needs given the various types of CWS in the Midwest.

The Panel addressed the four elements of the Agency's recommended framework separately below.

(1) Targeting Monitoring to the Most Vulnerable Areas - The Panel agreed that targeting the most vulnerable areas is appropriate. Caution was urged in how areas are defined as "vulnerable." One wants to insure that all CWSs with the potential for atrazine in their source water are included. The current method that uses an exceedance of 1.6 µg/L of atrazine in Safe Drinking Water Act (SDWA) samples as the criterion for inclusion of the system in the Atrazine Monitoring Program (AMP) is a good place to start. However, it was noted that the atrazine value from the SDWA sampling is an average of four quarterly values and it has been aptly demonstrated by EPA and the registrant that sampling at this frequency has the potential to underestimate the true atrazine concentration by a large amount. The use of the WARP model by EPA to determine if a CWS might be vulnerable to atrazine exceedances is an excellent second screen/trigger for inclusion in the monitoring program.

The Panel noted that agriculture is not a static enterprise and any monitoring program needs to have the flexibility to accommodate the large year to year variability in agricultural practices that results from market forces, regulation, changes in technology and other factors.

2) Targeting Intensive Sampling to the High Occurrence Period - The Panel agreed that intensively sampling during periods of expected high occurrence of exceedances is certainly appropriate. In explaining the occurrence of atrazine in surface waters over time, some Panel members suggested use of the following function:

$$\mathbf{Atrazine\_Occurrence_{surface\ water}(t) = Source\_Strength_{ATRAZINE}(t) + Hydrology(t)}$$

The two components are Source_Strength and Hydrology. Source_Strength relates to the amount of atrazine applied in a basin and that it is a function of when the application occurred in time. Hydrology refers to the fact that to move atrazine off site, water is needed, and this, too, is a function of time. From this equation, it is clear that intensive sampling of a runoff event before atrazine is applied in a basin is of limited

usefulness in describing Atrazine_Occurrence as would be intensive sampling after atrazine application when there has been no runoff to move atrazine into surface water.

Reservoirs need to be considered separately when designing a monitoring program. It can make a large difference in exposure whether the CWS uses a small run-off-river reservoir with a retention time of days, a large reservoir with a retention time of months to years, or an off-stream storage reservoir where water is pumped from the stream for later use by the water supply system. The USGS has done research on predicting the concentration of atrazine at the outflow of a number of reservoirs in the Midwest (Battaglin and Goolsby, 1998).

(3) Basing Sampling Frequency on the Toxicological Exposure Duration of Concern - The Panel believed that this approach would be appropriate if there were agreement among scientists on the toxicological exposure duration of concern. Some on the Panel expressed the opinion that there doesn't seem to be strong evidence of adverse health impacts from atrazine exposures at the levels found in surface waters; because of this, these Panel members believed it was unfair to ask the registrant to increase their sampling efforts. On the other hand, the fact that there is so much atrazine used in the United States, and that it is found in every environmental compartment (i.e., ground waters, surface waters, rainfall, lakes, reservoirs and some drinking water supplies) raises the importance of further monitoring. There may be subtle human health or environmental effects that currently remain unidentified or not well understood. The Panel believed that it is important to continue to monitor for atrazine in the environment, to better understand exposure levels and ensure that they continue to decline.

(4) Use Autosamplers to Collect Data for Exposure Periods of Interest - The Panel agreed that this is appropriate. Autosamplers currently are used successfully by many researchers who are studying the fate and transport of pesticides including atrazine. Autosamplers can better capture transient events that happen during periods when personnel are not available to do the sampling. Getting good reliable information from autosamplers also requires having good quality assurance and quality control programs in place.

EPA should look at other methods and procedures for collecting water samples. For instance, water treatment plant operators should be able to collect samples from their system's raw water intake and/or finished water outflow every day. These samples could be refrigerated or frozen with the decision to analyze any single or composite samples made at some later date.

Question 4.2

The April 2010 SAP recommended that simulations of candidate monitoring strategies and evaluations of the utility of different exposure estimation methods be benchmarked against intensive empirical data that cover a representative range of sites. In response, the USEPA proposes using more intensively sampled ambient water monitoring for

flowing waters and PRZM/EXAMS modeling for static water bodies, matching them to the CWS based on similar chemograph shapes (see Section 7.3).

a. Please comment on the Agency's proposal to use chemograph shapes (number, duration, and magnitude of peaks) to match CWS with more intensively monitored datasets. In particular, the Agency is interested in recommendations on approaches for matching chemograph shapes given the loss of short-duration peaks that occurs with less frequent sampling.

Panel Response

As noted in the question itself, the chemographs from the less frequently sampled sites may have short-duration concentration peaks that were not sampled. Matching a chemograph that is missing short-duration concentration peaks to chemographs from high frequency sampling sites will not show that short-duration concentration peaks are missing from the lower sampling frequency data. Instead, this approach will give a false confidence that the short-duration concentration peaks are not missing in the less frequently sampled data. In addition, chemograph shapes are influenced by factors such as pesticide use (amount and geographic distribution) and precipitation, as well as static watershed characteristics. A chemograph shape for a site can vary temporally because pesticide use and precipitation are temporally variable. Given the likelihood of missing short-duration concentration peaks and the variability of chemograph shapes over time at a site, using the chemograph shape as the mechanism to link CWS sites and the high sampling frequency sites may be overly difficult and may provide a false sense of confidence in terms of assessing the adequacy of sampling frequency for a site.

Functional Data Analysis (Ramsey and Silverman, 2002; 2005) may provide a potential approach to match sites based on chemograph shapes; however, care must be taken in accounting for missing peak concentrations when matching.

A suggested alternative approach is to match high frequency sampling sites to CWS sites based on water body and watershed characteristics. First, sites should be grouped by their source water type, such as static, flowing, or other. Within the source water groups, sites then can be further classified or grouped based on their watershed, pesticide-use, and climate (precipitation) characteristics, as well as their geographic locations. In terms of static or non-flowing waters, hydrologic characteristics such as retention time also should be used in this grouping exercise.

b. Please comment on the strengths and weaknesses of using more intensively sampled datasets from Heidelberg and AEEMP for analysis of flowing water and PRZM/EXAMS for lakes and reservoirs.

Panel Response

Appendix D very nicely shows the problems with trying to draw inference for an exposure when the interval is sampled only with a single point per interval. Sometimes one gets lucky with the timing of the sample, and sometimes one misses the peak

completely, thus grossly underestimating exposure. Thus, it is important to have true daily data when developing a methodology, as both the Heidelberg and AEEMP datasets do. PRZM/EXAMS modeling is also useful and provides calibrated curves with simulated daily values. While it should be noted that these models are not necessarily true data, and thus any conclusions drawn using them should be taken as approximations, they would be expected to give much more accurate results than a linear or stair-step interpolation of less frequently sampled data (such as weekly data). There is a substantial literature on the use of computer models to augment physical data, and when properly calibrated (i.e., adjusted to match the available sparser real data), the approximations can be very good (Kennedy *et al*., 2001; Higdon *et al*., 2004; Williams *et al*., 2006).

The Heidelberg dataset is high quality and represents frequent sampling, and it has the benefit of spanning multiple years, allowing for the understanding of some of the relationships with external driving conditions such as rainfall. However, there are definite limitations in its use. It does not represent geographical diversity (as the AEEMP datasets do) as its data are concentrated in a single region in Ohio. The stream characteristics (e.g., watershed parameters, pesticide usage patterns, climate) also are not necessarily representative of the broader range of flowing water sources in other parts of the country. It does not exhibit the scale effects that many water sources do. Generalization to other watersheds would need to be done with care to avoid incorrect predictions which could lead to a false sense of security in their accuracy.

The Perry Lake dataset would be more relevant for reservoirs. SEAWAVE was designed for flowing water, rather than for reservoirs, and PRZM/EXAMS for a somewhat different scenario. Furthermore, PRZM is a very detailed model, not the easiest to use in practice, and needs to be tailored to each individual watershed. SWAT is another model with similarities in design intent and need for specificity of application.

Each reservoir may be different, depending on drainage basin characteristics (e.g., size), capacity, turnover time, and other location-specific variables. A smaller stream-based reservoir may exhibit similar behavior to flowing water. However, a larger reservoir will generally have multiple input sources and slow turnover times, and thus pesticide concentrations will be much more smoothed (attenuated peaks) than for flowing streams. In such a case, sparser (e.g., weekly) sampling might be sufficiently representative of the overall chemograph, as demonstrated with the Perry Lake example. Working with flowing water could be seen as a conservative approach. Predicting peaks in reservoirs may be easier than for flowing water, since the peaks should be more broad and easier to catch with less frequent samples.

There was some sentiment among the Panel members that a more practical use of resources could be the collection of more varied, intensely-sampled datasets (particularly for reservoirs), rather than attempting to fit complex, labor-intensive, site-specific models to less-frequently sampled datasets.

Question 4.3

Once the magnitude and duration of exposure of toxicological concern are identified, the USEPA will determine which method(s) to use to analyze the existing atrazine CWS monitoring data for use in drinking water exposure estimates. Based on the April 2010 SAP recommendations, the Agency is focusing on regression-based modeling combined with random function modeling (for example, a revised WARP model coupled with kriging or the USGS SEAWAVE-Q model).

Please comment on the strengths and weaknesses of these proposed approaches and provide any recommendations for improving the model applications.

Panel Response

The easiest approach would be to have sampling twice during each period of interest. The current sampling scheme would be appropriate for a period of interest of 14 days or longer; otherwise, more frequent sampling would be recommended. Sampling could be manual or via autosamplers. This approach would give direct information about exposure over an interval, without the need for complex interpolation.

Combining a deterministic model, like SEAWAVE-Q or PRZM/EXAMS, with a regression-based model, like WARP, appears to be the most promising approach to deal with sparse data in flowing water (the scenario for which these models were designed). This approach makes use of information about the pathways and aggregate data that are not incorporated by purely statistical methods such as kriging or neural networks. However, the April 2010 SAP noted that, while exposure estimates generated through inference from other sites (such as WARP and other models) are appropriate for study design and screening-level evaluation, the uncertainty may be unacceptably high with these methods in terms of evaluating atrazine levels in order to implement regulatory actions. The purpose of models such as WARP is to provide concentration estimates and probabilities in order to guide more intensive monitoring and assessment. Similarly, WARPsim (Vecchia and Crawford, 2006) was developed to simulate daily pesticide concentrations based on WARP coupled with time-series data on precipitation and temperature, and includes a seasonal component to simulate the time-series concentrations. However, it was not intended to be used to predict actual concentrations for a specific day or location.

Combining a model such as WARP with a purely statistical method will give better interpolation than simple methods such as linear interpolation, but it would fail to make use of information about the expected shape of concentration curves that could be predicted by a deterministic model. However, there is a trade-off in the difficulty of implementing these methods, and it may be easier to implement a combination of WARP and a relatively simple statistical approach such as kriging, compared with combining WARP with a deterministic model. Such a consideration might have ramifications in the field.

Kriging does allow reasonable estimation of confidence bands, which could be useful for guessing what might be happening in between observations. (Please note that empirical variograms are highly variable). Slide 94 of the Agency's presentation clearly shows the need for joint estimation of variograms in order to get the confidence intervals right (one can get really wide confidence bands on the 16-day sampling). The correlation structure is clearly non-stationary and Gaussian may fit well overall, mainly because of the many small values. Around the peak(s), the structure is changing much more rapidly and does not look like a Gaussian correlation structure. Using a Gaussian correlation will seriously underestimate the peaks. One possibility to avoid fitting complex non-stationary models would be to use only the portions of the chemographs closer to the peak areas for fitting the variogram structure. While this may over-estimate variability in the low-level parts of the chemograph, those are generally not the regions of concern.

WARP has been developed as a nationwide model. In the case of the atrazine evaluation, it would be better to tailor it to the Corn Belt, which could require changes in the variables selected for inclusion in the model.

One Panelist noted that there is a need for a quality control plan. As currently there are no replicates collected, one doesn't have a good idea about sampling variability or measurement error. In the presence of both data and model uncertainty, one should be wary of overestimating the confidence level of any model predictions.

The Panel also discussed how this exercise will inform broader monitoring efforts around atrazine in the environment. At the April 2010 SAP, the Panel recommended the use of statistical models incorporating environmental explanatory variables and geostatistical correlations to simulate chemographs for input waters for CWSs. With such a system, the exposure end-points of interest can be estimated and proposed sampling procedures evaluated for their statistical properties. It is clear from the presentations made to this Panel that both EPA and the registrant continue to pursue this approach. It should be noted that it is possible to generate simpler empirical models of more complex simulation model output. These simpler models can be used in place of the more computationally intensive simulation models for rough predictions for larger geographic areas making national prediction possible. This model might be as simple as a linear regression, or more sophisticated like a Gaussian process emulator (Kennedy and O'Hagan, 2001; Santner *et al*., 2003; Gramacy *et al*., 2008).

It seems that a lot of time and effort is being concentrated on properly modeling the magnitude of concentrations in source (raw) waters. Realizing that source waters will eventually be modified through the CWS treatment process and that the ultimate exposure metrics for the human risk assessment are derived from the levels of atrazine in CWS output (drinking) waters, it is comforting to see that both EPA and the registrant are making advances in characterization of CWS water treatment processes as they affect atrazine concentrations.

As EPA and the registrant continue to explore sampling plans for monitoring, some thought should be given to actually using the simulation models and CWS

characterizations as part of the monitoring process. In particular, it is feasible that models will eventually be accurate enough to provide predictions of atrazine concentrations in source waters to a CWS for the coming crop season. Instead of requiring a CWS to collect and analyze water samples in their output (drinking water) stream at some pre-defined frequency, say daily or weekly in the case of some sites, it should be possible to use the models to facilitate targeting sampling to periods of time most likely to experience an exceedance.

Description of a possible sampling plan follows. Prediction of expected concentrations for raw and drinking water for a CWS would be run at the beginning of the crop year using available long-term climate predictions for the catchment area of the CWS, the expected cropping practices in this area, and CWS site-specific characteristics. It is not required that one know the exact days of rainfall or atrazine application; just having distributions for the expected number of heavy rainfall events and expected duration and likelihood that these events would coincide with new atrazine in the field would be sufficient data for this exercise. For most CWSs, the expectation would be that there would be no predicted exceedances. Sites with a high likelihood of exceedances might be required to implement interventions under specified conditions to mitigate expected exceedances. The CWS would be required to collect water samples on a periodic basis, for instance, daily, and store them in a form that would preserve atrazine concentrations. At the end of the season, the model would be run again, but this time with actual rainfall patterns and crop/ag chemical application/use information. The post-season model run would identify periods in the season just past where high atrazine concentrations would have been predicted. The stored water samples for this period would then be retrieved and analyzed to get actual atrazine concentrations in CWS output waters. This information could then be used to assess CWS performance and the need for further system management changes to ensure atrazine concentrations in the future are below maximum allowable concentrations.

The described plan closely follows sampling protocols that are typically used to obtaining information on rare events. A benefit is that sampling effort can be more easily controlled and planned. CWS managers would know that they would be responsible for only X water sample analyses in any one year, but would not know exactly which time periods would be analyzed. They would be responsible for collecting and storing water samples for the season. During this meeting's public comment period, the Panel heard that the economic impact of the sampling frequencies required is of high concern for CWS managers. This plan would control the sampling effort and, at the same time, focus sampling on periods with the highest likelihood of actually detecting measurable concentrations. It acknowledges the fact that for most CWSs and, for most times, there is little to no detectable atrazine concentration in output waters. It also allows for random selection and analysis of stored water samples to estimate the extent to which models are missing other peak periods, if any.

It is recognized that this method of organizing sampling effort deviates from the traditional approach which specifies that the sampling effort be distributed uniformly over the time period of interest. However, it has been made clear in the presentations to

the Panel that traditional thinking will lead to intensive and expensive sampling at high frequency with the expectation that the vast majority of analyses will yield "no detect" concentrations. While this may provide confidence to many that there is little or no atrazine in CWS output waters, this is an unnecessarily expensive approach that is not likely to provide better results than the more focused, less expensive sampling plan suggested here.

Question 5.0: Scientific Considerations in Potential Sensitivity of Infants & Children

As described in Section 6.0 in the draft Issue Paper, the FFDCA, as amended by the FQPA (1996), requires the Agency to give special attention to infants and children by placing emphasis on the availability of toxicology and exposure information to estimate the potential risk to these age groups. The 2003 Registration Eligibility Decision (RED) notes that the Agency reduced the FQPA Safety Factor for those localities which are part of the drinking water monitoring program conducted by Syngenta, as required by EPA. This reduction in the Safety Factor was based on more intensive monitoring required in the monitoring program that reduced the uncertainty for estimating concentrations of atrazine and its metabolites in drinking water for the 90-day rolling average identified in the previous risk assessment as the appropriate duration of concern. Based on the newest studies, our understanding of atrazine's temporal pharmacokinetic and pharmacodynamic characteristics has changed and it appears that shorter durations may be appropriate. In light of the potential need to shorten the critical duration of exposure and as discussed by the Panel in Questions 4.1-4.3, the Agency is evaluating appropriate methods for assessing the drinking water monitoring data.

Previously, the Agency noted data gaps related to "information on Atrazine concerning exposure throughout all critical developmental periods (i.e., gestation through puberty in both sexes), in particular, early in development…. (emphasis added, FQPA memo, 2002)." At the time of the previous risk assessment, there were only a few available studies which included measures sensitive to endocrine disruption on specific early life exposure periods (e.g., peripubertal). Since that time new studies have become available (Section 6.0) which add to the existing database and support the findings of the older studies. In addition, new tissue dosimetry studies provide information on the transfer of atrazine and its chlorinated metabolites to the fetus and to the lactating pup. Further, a study evaluating exposure to atrazine from multiple critical developmental periods identified previously by the Agency as a data gap is on-going. The results of this research will inform our understanding of the potential for differential life stage sensitivity.

In the coming months, as the drinking water exposure analysis develops further and the on-going toxicology studies become available, the Agency will work towards completing the scientific analysis that will inform whether or not a new FQPA Safety Factor should be applied in the atrazine risk assessment.

The Agency requests the Panel to comment on important scientific factors for the Agency to consider in its analysis. Please include in your comments specific consideration of

uncertainties in estimating drinking water exposures and remaining uncertainties in atrazine's toxicological profile across life stages, particularly as they pertain to assessing risk to infants and children.

Panel Response

Consideration of the value of the FQPA safety factor is seemingly best predicated on transparent and systematic consideration of the most important qualitative and quantitative uncertainties associated with both exposure and effect, relevant to susceptible life stages, in a context consistent with that for other pesticides. In view of the fact that the database for atrazine relevant to the selection of this factor is still evolving, reference here is to some of the generic aspects that might be considered explicitly based on outcome of additional analysis, including the following:

1. For exposure, this could relate to the likelihood of capturing the relevant period(s) of susceptibility or over- or underestimating exposure for all life stages, with the proposed monitoring strategy (including, for example, consideration of determination of total chlorinated triazines (TCT) rather than atrazine alone).

2. For effect, some critical questions and/or aspects to be addressed in this context include the following:

a. To what extent does the database on hazard and kinetics inform us about the potential for increased susceptibility of infants and children?
b. Is the early key event or late adverse effect for the critical effect sufficiently protective for all age groups, based on hazard characterization (including knowledge of mode of action)?
c. How protective is it (e.g., is an early key event protective for later adverse effects)?
d. What is the impact of the potential reliance on a benchmark dose (versus an empirically-observed effect/no effect level) in relation to uncertainty in the characterization of the relevant dose-response relationship?
e. Does the degree of conservatism associated with use of the lower confidence limit for a benchmark dose increase confidence?
f. While the epidemiological data are not considered sufficiently robust for inclusion in a quantitative risk assessment at this time, can data from any of the studies that are considered of highest quality be used to provide some idea of relative sensitivity of various age groups of the human population?

Even though considerable information related to water monitoring for atrazine is available, the overall potential for dietary intake of atrazine in young children or adolescents remains somewhat uncertain. It is recognized that information on atrazine and its degradates in source water is valuable. Certain water treatment technologies can reduce or eliminate atrazine; thus, finished drinking water data should be used in the determination of exposure via this source. The existing risk assessment for atrazine uses a 90 day rolling average to estimate exposures. The Agency suggested that better

estimation of the duration of dosing most relevant for inducing developmental toxicity would be useful, but it remains unclear at the present time what the relevant duration of dosing for either adults or developing organisms may be.

There is a general consensus from the results of the animal studies that atrazine can influence the GnRH-mediated LH surge following oral exposures leading to reproductive and/or developmental toxicity. The 2003 IRED (USEPA, 2003) for atrazine uses the 6-month dietary atrazine study (Morseth *et al*., 1998) as the critical study for establishing an intermittent/chronic RfD. Recent studies submitted by the registrant (Coder *et al*., 2010) suggest that prolonged dietary exposures to atrazine do not influence the LH surge. Such qualitative differences in response in studies using relatively similar dosing strategies is difficult to reconcile, although some factors (e.g., different strains of rats used) may contribute to the conflicting findings.

Toxicokinetic studies using radiolabel indicate that the levels of atrazine and its metabolites in the fetuses of treated dams are almost equal to those measured in the blood of the dams, suggesting that the placenta offers little barrier for these chemicals (Stoker *et al.*, 2010; Kamel *et al.*, 2010). On the other hand, the data show differential kinetics in the adult versus the pup during lactation with a significant drop-off in the concentration of atrazine and its metabolites as the chemicals move from the dam‘s mammary gland to milk in the pup stomach, to pup plasma and pup brain, such that the concentrations in the pup plasma and brain are approximately 10-fold (or more) less than that in the dam plasma (Stoker and Cooper, 2007; Stoker *et al.*, 2010; Kamel *et al.*, 2010). One of the most important uncertainties is the magnitude of the internal dose in the adults and the fetuses/pups. The newer kinetic studies are a step in the right direction for characterizing these values. The feeding regimen is a more realistic paradigm for delivering the dose, i.e., more reflective of periodic drinking water exposure. The indication from the data presented that a pseudo-steady state is occurring with repeated exposures is useful information for internal dose estimation and probably needs more verification. An extension of the need for data on internal dose is that it is needed for developing a PBPK model that will better estimate the internal concentrations in people. This activity seems to be underway and a PBPK model should be forthcoming. Some of the kinetic parameters for atrazine metabolism need to be determined for infants and children. While certainly a start, more than one human hepatocyte sample needs to be evaluated to develop metabolic parameters for the PBPK model.

The findings generally suggest, however, that there may not be selectively higher exposures to the developing organism during the prenatal or lactational period. A number of experimental studies have evaluated toxicity following prenatal, postnatal or peripubertal exposures to atrazine. With prenatal exposures (Rayner *et al*., 2005; Rosenberg *et al*., 2008), effects on endpoints including preputial separation and mammary gland development were noted with repeated exposures of 50-100 mg/kg/day. With combined prenatal/lactational exposures (Rayner *et al*., 2007), effects on preputial separation and prostatitis were observed with repeated 100 mg/kg/day exposures. Several studies evaluated effects of atrazine with peripubertal exposures (e.g., PND22-44, Stoker *et al*., 2000; Laws *et al*., 2000; Trentacoste *et al*., 2001; Friedmann *et al*., 2002; Pogrmic

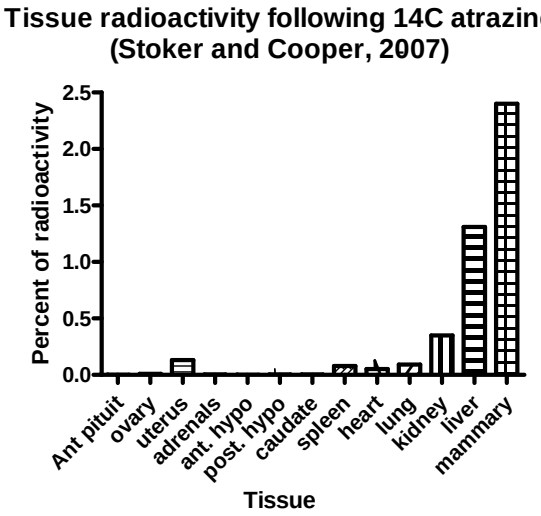
et al., 2009). Preputial separation was affected at dosages as low as 12.5 mg/kg/day while testosterone levels were reduced at higher exposures (50 mg/kg/day). Thus, none of these studies indicate that the prenatally, lactationally, or peripubertally-exposed animal exhibits higher sensitivity to developmental disruption than the effects noted in adults related to changes in LH surge.

The Agency has noted in its recent atrazine assessments (i.e., the 2003 IRED and 2006 RED) that the database available to support decision-making with regard to the FQPA 10X Safety Factor was not adequate to provide the necessary understanding of the potential for pre- and postnatal toxicity and/or exposure during these life stages, such that the Safety Factor could be reduced or removed in all settings. A reduction of the Safety Factor was deemed appropriate for assessments of those sites where the mandatory sampling survey was being conducted, in the belief that exposure to this subpopulation was being satisfactorily characterized. This decision will be revisited in the future, once the toxicology-related questions and related monitoring frequency issues are resolved. It remains unclear, however, whether existing data coupled with as yet unfinished studies will answer sufficiently the questions about the toxicity profile for all important life stages (gestational, lactational and peri-pubertal) and whether there are any life stage-related differences in sensitivity.

One of the biggest uncertainties at the present time is the validity of the very low dose effects observed on mammary gland development in the Enoch *et al*. (2007) paper. The data suggest that alterations in development occur following repeated exposures to a mixture of atrazine and metabolites/degradates of atrazine (0.09-8.7 mg/kg/day from GD15-19; mixture of hydroxyatrazine (ring hydroxylation), deethyl atrazine, deisopropyl atrazine and diaminochloroatrazine). Certainly if these very low dose effects are real, they will need to be considered as representing a highly sensitive endpoint for risk assessment. However, because these mammary gland effects occurred at doses so much lower than have been observed for other endpoints, these data need to be replicated. These data arose from exposure to a mixture and the other data arose from atrazine treatment alone. However, if these data are reproduced in a new study, then further investigation of the effects of the mixture components will be warranted.

Several questions concerning the data in Enoch *et al*. (2007) were noted by the Panel. There was some continuing uncertainty about the scoring of histological lesions across studies. The ranked data appeared inappropriately evaluated using parametric analyses, although averaging values for pups within a group before analysis may alleviate those concerns. There were low mean scores in one of the control groups (PND33). There was generally little change in effect at higher dosages, although with later time-points (PND40 and 60) the general tendency was more extensive reduction in mammary gland development scores with higher dosing. It is reasonable, however, to assume that, even if the results do not show clear dose-related responses, a developmental alteration could be an all or none response across exposure levels, once a minimal threshold exposure was passed.

One interesting piece of information that did not come out in the EPA presentations during the meeting was that the mammary gland accumulates considerable radioactivity following oral administration of radiolabeled atrazine (Stoker and Cooper, 2007). The figure below taken from Stoker and Cooper (2007) shows the percentage of total radioactivity in various tissues three hours following oral administration of radiolabeled atrazine. Note the high concentration in the dam's mammary glands compared with that in a variety of other tissues.



It should be stressed that earlier studies also reported mammary gland effects with early in life atrazine exposure (Rayner *et al*., 2005), but with markedly higher levels of exposures. The study reported by Enoch *et al*. (2007) is the only study evaluating mixtures of atrazine and its degradation products. Thus, while no other studies have used a mixture of atrazine and degradation products to study developmental effects and those results need to be confirmed in further studies, there is some evidence to suggest that the mammary gland may be both highly exposed and sensitive to disruption by early atrazine exposures.

The inclusion of the environmental degradate 4-hydroxyatrazine (ring hydroxylation) in the Enoch *et al*. (2007) study may indicate that this particular contaminant is worthy of some focused evaluation. While not a product of mammalian metabolism of atrazine, 4-hydroxyatrazine is a substantial environmental contaminant resulting from atrazine degradation (approximately 20% of the sample used in the AMM) and it may warrant further study. There is a substantial number of toxicity studies on 4-hydroxyatrazine that were not included in the materials made available to the Panel for the meeting. As for the parent atrazine and its chlorinated metabolites, data exist for 4-hydroxyatrazine which evaluate the potential for genotoxicity/mutagenicity, acute toxicity, short-term, subchronic and chronic repeated oral dosing toxicity, developmental toxicity and carcinogenicity. EPA has established an RfD of 0.01 mg/kg/day for 4-hydroxyatrazine, based upon the observation of histopathological lesions in the rat kidney in the chronic/carcinogenicity study (USEPA, 2006). The 2007 meeting of the FAO/WHO Joint Meeting on Pesticide Residues derived an acceptable daily intake (ADI)

of 0.04 mg/kg/day for this degradate, also based upon the renal toxicity observed in the same study (FAO/WHO JMPR, 2008). Furthermore, preliminary data suggest that 4-hydroxyatrazine does not possess the neuroendocrine properties observed with atrazine and its chlorinated metabolites, i.e., up to a 400 ppm dietary exposure to 4-hydroxyatrazine did not influence the LH surge (Eldridge *et al*., 2001). Based on these studies, 4-hydroxyatrazine does not appear to pose an additional hazard related to atrazine.

The very low exposure levels (0.09-8.7 mg/kg/day) of the mixture suggest that mammary gland effects may be sensitive endpoints following gestational exposures to atrazine along with degradation products (Enoch *et al*., 2007). While the mechanism for the non-cancer mammary gland effects following atrazine/degradate exposures is unknown, these findings along with the apparently substantial accumulation of atrazine and/or its metabolites in the mammary gland provide some concern for higher sensitivity in developing organisms. Replication and extension (e.g., dosimetry, 4-hydroxyatrazine and other individual degradation product effects) of findings on altered mammary gland at low level exposures are necessary, however, to inform the quantitative risk assessment. If the Enoch *et al*. (2007) studies are replicated, subsequent studies should specifically evaluate the role 4-hydroxyatrazine may play, either alone or in environmentally-relevant mixtures, in the alteration of mammary gland development. There may also be a need to consider direct dosing of pups during lactation to determine the potential for (differential) sensitivity in that population.

The (re)-consideration of the FQPA 10X Safety Factor should start with documentation and conclusions regarding the adequacy of the data describing atrazine's toxicity profile in adults in order to make intelligent conclusions regarding comparative sensitivities. Once that is completed, the data for earlier life stages can then be presented and more clearly interpreted.

Question 6.0: Implications of MOA & Toxicity Profile on Water Monitoring

As described in the Introduction of the draft issue Paper (Section 1), the goals of the current atrazine re-evaluation are: 1) to determine the extent to which new science indicates a need for the Agency to develop a revised human health risk assessment for atrazine and 2) to re-consider, as appropriate, the frequency of drinking water monitoring needed. Two important issues related to achieving these goals are determining the point of departure and identifying the critical effect and its associated duration of exposure. Proposed updates to the point of departure are considered in Questions 3.1 and 3.2. With respect to the critical duration of concern, the frequency of drinking water monitoring is related to the temporal pattern of the toxicological endpoint of concern used for the risk assessment. Generally, longer durations of toxicological concern (e.g., a long-term chronic effect) require a less frequent drinking water sampling design to approximate longer term exposures. However, as the duration of concern shortens, the frequency and timing of sampling become more important in determining how well the sample data capture short-duration exposures. Observation epidemiology studies raise hypotheses and suggest possible links between atrazine exposure and

reproductive and developmental outcomes, but these epidemiology studies suffer from limitations which prevent firm conclusions. Although certain studies some provide qualitative support for the human relevance of the endpoints identified through the experimental toxicological database, they provide little to no information on the critical duration of exposure. In addition, the MOA and PK database are also lacking in human specific information on the effects of atrazine which could be used to quantitatively extrapolate between species. As such, the information available to evaluate the critical duration of exposure lacks quantitative precision. Thus, the critical duration of exposure is instead derived by inferring generic knowledge from a variety of scientific disciplines.

The Agency has used a variety of approaches to extrapolate findings from experimental animal data to humans including allometric scaling and human-specific information on the physiology of the menstrual cycle inferred from the IVF literature. According to the Agency's analysis of the pharmaceutical data and allometric scaling of the rodent pharmacokinetics data, the potential durations of human exposure that would correspond to the exposure period of interest in rats lie between a few days to approximately 4 weeks of exposure.

Please comment on the Agency's analysis and preliminary conclusions contained in Section 8.0 of the draft Issue Paper as it relates to the potential critical windows of exposure. Please include in your comments additional or alternative approaches or data that may inform this issue.

Panel Response

Section 8 in the Issue Paper systematically addresses 1) the use of LH surge suppression in the rat as the BMR, 2) the strength of evidence linking the BMR to a variety of endpoints identified in the human epidemiology and experimental animal studies, 3) the comparative timing and extent of exposures with respect to the potential to suppress the LH surge in animals and humans, including kinetic and dynamic considerations, 4) the potential for water atrazine concentrations exceeding a level of concern to be missed by current sampling procedures, and 5) the concept that sampling frequency can be meaningfully adjusted based on the potential for human health outcomes. The Agency has done a good job of summarizing the situation in each of these areas with respect to the uncertainties and limitations in both the data and in our scientific understanding of what the data are telling us.

The Agency has determined that the collected information suggests an optimal water sampling frequency would be between a few days and four weeks based on durations of exposure considered relevant with respect to potential human health outcomes. Currently, the sampling frequency required of the registrant is once a week during the use season and once every two weeks during the rest of the year. Several Panel members indicated that they felt this sampling frequency was adequate. No one raised strong objections or could offer a rationale for an alternative sampling frequency based on the collected information discussed during the SAP meeting.

A series of assumptions and extrapolations contribute to the proposed critical window of human exposure and given the collective uncertainties that these assumptions introduce, the imprecision in the Agency's proposed sampling frequency seems justified. This may be about as precise an estimate as can be obtained when starting with the experimental animal data and the exposure requirements for LH surge suppression as opposed to using outcomes that are more unequivocally adverse. In this regard, the consideration by the Agency of the human relevance of the adversity of LH surge suppression on the basis of both pharmacokinetics and pharmacodynamics, and taking into account the broader database including that on pharmaceutical agents used to block the LH surge, is to be commended. However, one Panel member commented that it was difficult to assess the relevance of the information from the GnRH antagonist research since a different mode of action was involved and no data were provided on the minimum concentration of atrazine that would be needed over the period of exposure that would produce an equivalent reduction in GnRH level to produce a similar effect at the critical period.

One question that clearly needs further consideration is whether there is a critical AUC (i.e., exposure of a target site) that leads to a given level of suppression of the LH surge. This would be tenable if the rat estrous cycle was equally susceptible over all of the four days, but not if tolerance changed over the critical period (for instance, if an AUC of 100 units on day one was sufficient, but only 50 was necessary on day 2). There appears to be a critical exposure time, but is this 2 days, 3 days, or 4 days? Indeed there could be a critical exposure time combined with a critical level. Although this would have the same units as AUC it would have a minimum exposure time (possibly a fixed exposure time). In order to be sure that AUC is related to biological response, it would be necessary to obtain the same AUC by using different dosing regimens (e.g., one small dose followed by a large dose followed by two small doses, or four moderate doses giving same AUC) but not necessarily by reaching a pseudo-steady state. If the different patterns of dosing give the same toxicological endpoint, then it is AUC that is the important measure of exposure. However, even if one could sort through this, and apply this information to estimates of the proposed range of human exposures responsible for a potential adverse outcome, the answer would still be little more than an educated guess.

Question 6 specifically requests alternative approaches. There is another way of approaching this that may be useful at least when setting the boundaries on exposures that may present a concern for human health effects. The current epidemiology database is characterized as providing suggestive evidence that the mechanisms of action thought to be operative in rats may be occurring in humans exposed to atrazine. The Agency has appropriately concluded that the limited human evidence is insufficient to establish causality and does not provide sufficient quantitative exposure information to use in a risk assessment. However, what if one assumes that the reported human health outcomes are, in fact, due to current levels of exposure to atrazine? Although the water sampling data may not be adequate to assure that atrazine peaks are captured in all water systems, clearly some of the patterns seen are based on rather comprehensive data sets. These patterns of atrazine concentrations in water could be used to provide reasonable estimates of the extent and duration of human consumption of atrazine following agricultural

applications for pre-emergent weed control. Simple models could be used to estimate human exposures corresponding to a range of times of exposure to the elevated concentrations observed in the field, given the expected maximum water consumption. Where spikes in water concentration are short lived and concentrations of atrazine change over a short time scale, then steady state blood levels will not be achieved. In this instance, one might consider what would be average levels of exposure, relate this to the equivalent internal exposure in the rat, and use the promised PBPK model to extrapolate to humans.

This may represent a reasonable alternative approach to determining levels of atrazine in drinking water that may represent risks to human health. These risks could be compared, certainly on an order of magnitude scale, with those calculated from the animal data, and may provide a lower bound conservative floor from which to work, and provide a different perspective on the water sampling frequency problem. This would put the Agency in a much better position if, in fact, the Agricultural Health Study or other epidemiology studies provide further support for human health effects as the results continue to accumulate.

The other consideration when faced with the uncertainty over a critical exposure period of from a few days to 4 weeks is whether basing sampling frequency on human health effects is, in fact, the best course of action. As was discussed in greater detail in the Panel's responses to Question #4, atrazine concentrations in source water and, subsequently, in the finished water supply of a CWS are dependent upon many factors that vary spatially and temporally. As noted, each CWS is unique in factors that affect the delivery of atrazine such as drainage basin size, characteristics of the soils, cropping pattern, slope, and whether the CWS source water intake is directly in the stream, a reservoir, or an off stream storage facility. In addition, there are many factors that affect the ability of a water treatment system to remove atrazine from water, such as the use of activated carbon and the type of oxidant used for treatment. It also has been shown that the amount of atrazine in a system can sometimes be related to the ongoing maintenance of the treatment plant. It may be more useful to consider a strategy that attempts to capture the pattern of atrazine in the source water of each CWS based on the characteristics of that particular water system, as opposed to a one size fits all approach based on the series of health-based considerations put forth by the Agency. Given the collective limitations of the health outcome-based approach, this would seem prudent, and would again put the Agency in a better position to take further action should the results of ongoing or future epidemiology studies prove more convincing.

In the meantime, since water is the primary source of environmental exposures to atrazine and its metabolites/degradates, there would be value in doing a better job of establishing the relationship between the measured concentrations at the community level, and the resulting absorbed dose in humans. This question could be first addressed using databases such as NHANES and could be supplemented with individual biomonitoring studies to determine the correlations between water concentrations pre- and/or post-treatment, residential tap samples and human urinary metabolites.

To conduct well designed epidemiologic studies in individuals in the future, additional individual-specific measures would be necessary. Important measures of water consumption and factors that modify exposures would include the following: measures of water consumption from the home tap, whether the water is filtered, how long the water is run from the tap before taking a drink (known to be important for metals), bottled water consumption (very important for women), consumption at work and at home (different communities or sources of water), weeks spent traveling away from home (i.e., drinking from a different community source), water consumption from food (i.e., canned tomatoes) and other bottled beverages.

Technical comments - On page 128 and in Table 8.1 of the Issue Paper EPA estimates human equivalent doses (HED) using an allometric scaling approach. The dosimetric adjustment factor incorporates an average female body weight of 60 kg (~132 pounds). However, the average body weight for a female in the U.S. has not been that low since about 1960, when it was 140 pounds (Ogden, *et al*., 2004). The most recent data published by the National Center for Health Statistics covers the period of 2003-2006 (McDowell, *et al*., 2008). The mean body weight for an adult female > 20 years old is now up to 164.7 pounds or ~75 kg. It is recommended that if the Agency chooses to go forward with an allometric scaling approach as a method for estimating a human equivalent dose that it employ the most current value for the average adult female body weight.

One Panel member questioned the appropriateness of using body surface scaling to compute an HED, given that atrazine in the parental form is not hypothesized to be the principal/sole toxic moiety. The adjustment of clearance based on body surface area adjustment is frequently applied to ensure equal blood concentration of parent chemicals in both species; this approach cannot directly be used to compute an HED from the rat on the basis of total triazine-equivalents (page 128, paragraph 2). In other words the use of an allometric dose adjustment factor to calculate an HED would appear to contradict the use of AUC for triazine-equivalents proposed to be used in the dose-response analysis (Chapter 3 of the Issue Paper).

REFERENCES

Alvarez-Sanz, M.C., Liu, J.M., Huang, H.H. and E.J. Hawrylewicz . (1986). Effect of dietary protein on morphologic development of rat mammary gland. J Natl Cancer Inst. 77(2):477-87.

Battaglin, W. A. and D. A. Goolsby. (1998). Regression models of herbicide concentrations in outflow from reservoirs in the Midwestern USA, 1992–1993. J Amer Water Res Assoc 34:1369–1390.

Boettger-Tong, H., Murthy, L., Chiappetta, C., Kirkland, J.L., Goodwin, B., Adlercreutz, H., Stancel, G.M., and S. Mäkelä. (1998). A case of laboratory animal feed with high estrogenic activity and its impact on *in vivo* responses to exogenously administered estrogens. Environ. Health Perspect. 106(7):369-373.

Breckenridge, C.B., Kim, D. and T.P. Pastoor. (2010). Atrazine: Preliminary results from new studies on atrazine. Report Number TK0040016. Syngenta Crop Protection, Inc. Greensboro, NC

Bromberger, J.T., K.A. Matthews, L.H. Kuller, R.R. Wing, E.N. Meilahn, and P. Plantinga. (1997). Prospective Study of the Determinants of Age at Menopause. Am J or Epidemiol. 145(2): 124-133.

Brown, N.M. and D.R. Setchell. (2001). Animal models impacted by phytoestrogens in commercial chow: implications for pathways influenced by hormones. Lab. Invest. 81(5):735-747.

Burger, H.G., Dudley, E.C., Hopper, J.L., Shelley, J.M., Green, A., Smith, A., Dennerstein, L., Morse, C. (1995). The endocrinology of the menopausal transition: a cross-sectional study of a population-based sample. J. Clin. Endocrinol. Metab. 80:3537-3545.

Coupe, R.H., and J.D. Blomquist. (2004). Water-soluble pesticides in finished water of community water supplies: AWWA 96(10):56-68.

Dooley G.P., K.F. Reardon, J.E. Prenni, *et al*. (2008). Proteomic analysis of diaminochlorotriazine adducts in Wistar rat pituitary glands and Lbeta T2 rat pituitary cells. Chem Res Toxicol. 21:844-851.

Dooley, G.P., Ashleya, A.K., Legare, M. E., Handa, R.J. and W.H. Hanneman. (2010). Proteomic analysis of diaminochlorotriazine (DACT) adducts in three brain regions of Wistar Rats, Toxicol. Lett. 199(1):17-21.

Eason, R.R., Velarde, M.C., Chatman, L. Jr, Til, S.R., Geng, Y., Ferguson M., Badger, T.M. and R.C. Simmen. (2004). Dietary exposure to whey proteins alters rat mammary gland proliferation, apoptosis, and gene expression during postnatal development. J Nutr. 134(12):3370-7.

Eldridge, J.C., Minnema, D., Breckenridge, C.B., McFarland, J. and J.T. Stevens. (2001). The effect of 6 months feeding of atrazine or hydroxyatrazine on the luteinizing hormone surge in female Sprague Dawley and Fisher-344 rats. The Toxicologist. Abstract 1525.

Fakhouri, W.D., Nuñez, J.L.and F.Trail. (2010). Atrazine binds to the growth hormone-releasing hormone receptor and affects growth hormone gene expression. Environ Health Perspect. 118(10):1400-5.

FAO/WHO JMPR. (2008). Pesticide residues in food — 2007. Toxicological evaluations World Health Organization, Geneva, Switzerland. pp.37-138.

Gold, E.B., Eskenazi, B., Hammond, SK., *et al*. (1995). Prospectively assessed menstrual cycle characteristics in female wafer-fabrication and nonfabrication semiconductor employees. Am J. Ind. Med. 28:799-816.

Gold, E.B., Bromberger, J., Crawford, S., *et al*. (2001). Factors associated with age at menopause in a multi-ethnic population of women. Am J Epidemiol 153:865-874.

Gold EB, Blount BK, Rasor MO, Lee, JWS, Srivastav A, Kim K. Thyroid hormones and thyroid disease in women in relation residents living near a Superfund site. Submitted 2010.

Gramacy, R. B. and H.K.H. Lee. (2008) "Bayesian treed Gaussian process models with an application to computer modeling." Journal of the American Statistical Association. 103:1119-1130.

Harvey, P.W, and P.F. Chevins. (1987). Crowding during pregnancy delays puberty and alters estrous cycles of female offspring in mice. Experientia. 43(3):306-8.

Higdon, D., Kennedy, M., Cavendish, J., Cafeo, J., and R.D. Ryne. (2004). Combining field observations and simulations for calibration and prediction. SIAM Journal of Scientific Computing. 26:448-466.

Hilakivi-Clarke, L., Onojafe, I., Raygada, M., Cho, E., Clarke, R. and M.E. Lippman. (1996) Breast cancer risk in rats fed a diet high in n-6 polyunsaturated fatty acids during pregnancy. J Natl Cancer Inst. 88(24):1821-7.

Hilakivi-Clarke, L., Stoica, A., Raygada, M. and M.B. Martin. (1998). Consumption of a high-fat diet alters estrogen receptor content, protein kinase C activity, and mammary gland morphology in virgin and pregnant mice and female offspring. Cancer Res. 58(4):654-60.

Joo, H., Choi, K. and E. Hodgson. (2010) Human metabolism of atrazine., Pestic. Biochem. Physiol., 98:73–79 (In press).

Jori, A., Pescador, R. and C. Pugliatti. (1971) Rat strain differences in the activity of hepatic microsomal enzymes. Biochem Pharmacol 20(10):2695-701.

Kakizaki, M., Kuriyama, S., Sone, T., Ohmori-Matsuda, K., Hozawa, A., Nakaya, N., Fukudo, S. and I. Tsuji (2008). Sleep duration and the risk of breast cancer: the Ohsaki Cohort Study. Br J Cancer. 99(9):1502-5.

Kennedy, M. and A. O'Hagan (2001) Bayesian Calibration of Computer Models. J R Stat Soc, Series B. Stat Methodol. 63:425-464.

Kingsbury, J.A., G.C. Delzer and J.A. Hopple (2008). Anthropogenic organic compounds in source water of nine community water systems that withdraw from streams, 2002–05: U.S. Geological Survey Scientific Investigations Report 2008–5208, 66 p. http://pubs.usgs.gov/sir/2008/5208/.

Kirchick HJ, Keyes PL and B.E. Frye (1978) Etiology of anovulation in the immature alloxan-diabetic rat treated with pregnant mare's serum gonadotropin: absence of the preovulatory luteinizing hormone surge. Endocrinol. 102(6):1867-73.

Klein, S.L. and D.R. Rager (1995). Prenatal stress alters immune function in the offspring of rats. Dev Psychobiol. 28(6):321-36.

Krieg, R.J., Tokieda, K., Chan, J.C. and J.D. Veldhuis (2000) Impact of uremia on female reproductive cyclicity, ovulation, and luteinizing hormone in the rat. Kidney Int. 8(2):569-74.

Lehrer, S. (1986). Rats on 22.5-hr light:dark cycles have vaginal opening earlier than rats on 26-hr light:dark cycles. J. Pineal Res. 3(4):375-8.

Lin, E.Y., Gouon-Evans, V., Nguyen, A.V. and J.W. Pollard. (2002). The macrophage growth factor CSF-1 in mammary gland development and tumor progression. J Mammary Gland Biol Neoplasia. 7(2):147-62.

MacLusky, N.J., Naftolin, F. and C. Leranth. (1988). Immunocytochemical evidence for direct synaptic connections between corticotrophin-releasing factor (CRF) and gonadotrophin-releasing hormone (GnRH)-containing neurons in the preoptic area of the rat. Brain Res. 439(1-2):391-5.

McDowell, M.A., Fryar, C.D., Ogden, C.L. and K.M. Flegal. (2008). Anthropometric reference data for children and adults: United States, 2003-2006. National Health Statistics Report No. 10. (PHS) 2009-1250. October 24, 2008. 45 pp. http://www.cdc.gov/nchs/data/nhsr/nhsr010.pdf.

Mediavilla, M.D., San Martín, M. and E.J. Sánchez-Barceló. (1992). Melatonin inhibits mammary gland development in female mice. J Pineal Res. 13(1):13-9.

Muhlhauser, A., Susiarjo, M., Rubio, C., Griswold, J., Gorence, G., Hassold, T., and P.A. Hunt. (2009). Bisphenol A effects on the growing mouse oocyte are influenced by diet. Biol Reprod. 80:1066-1071.

Ogden, C.L., Fryar, C.D., Carroll, M.D. and K.M. Flegal. (2004). Mean body weight, height, and body mass index, United States 1960--2002. Advance data from vital and health statistics; no 347. Hyattsville, MD: US Department of Health and Human Services, CDC, National Center for Health Statistics. http://www.cdc.gov/nchs/data/ad/ad347.pdf.

Owens, W., Ashby, J., Odum, J. and L. Onyon. (2003). The OECD program to validate the rat uterotrophic bioassay. Phase 2: dietary phytoestrogen analyses. Environ Health Perspect. 111:559-567.

Plagemann, A. (2005). Perinatal programming and functional teratogenesis: impact on body weight regulation and obesity. Physiol Behav. 15:661-8.

Ramsey, J.O. and B.W. Silverman. (2002). Applied Functional Data Analysis, Second Edition. Springer-Verlag, New York, NY.

Ramsey, J.O. and B.W. Silverman. (2005). Functional Data Analysis, Second Edition. Springer-Verlag, New York, NY.

Santner, T. J., Williams, B. J., and W.I. Notz. (2003) The Design and Analysis of Computer Experiments. Springer-Verlag, New York, NY.

SAP. (2010). FIFRA Scientific Advisory Panel (FIFRA SAP). A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding: Re- Evaluation of Human Health Effects of Atrazine: Review of Experimental Animal and *In Vitro* Studies and Drinking Water Monitoring Frequency. SAP Minutes No. 2010-04. FIFRA Scientific Advisory Panel, Office of Science Coordination and Policy, Office of Prevention, Pesticides and Toxic Substances, U.S. Environmental Protection Agency. Washington, DC. Available: http://www.epa.gov/scipoly/sap/meetings/2010/042610meeting.html

Sowers, M.F., Crawford S., Morgenstein D., *et al*. (2000) Design, survey sampling and recruitment methods of SWAN: A multi-center, multi-ethnic, community-based cohort study of women and the menopausal transition. Lobos R, Marcus R, Kelsey JL (eds.), Menopause. New York: Academic Press, p. 175-188.

Stanko, J.P., Enoch, R.R., Rayner, J.L., Davis, C.C., Wolf, D.C., Malarkey, D.E., and S.E. Fenton (2010) Effects of prenatal exposure to a low dose atrazine metabolite mixture on pubertal timing and prostate development of male Long-Evans rats. Reprod Toxicol (2010), doi:10.1016/j.reprotox.2010.07.006. In press.

Tamarkin, L., Cohen, M., Roselle, D., Reicher, C., Lippman, M. and B. Chabner. (1981). Melatonin inhibition and pinealectomy enhancement of 7,12-dimethyl-benz(a)anthracene-induced mammary tumors in the rat. Cancer Res. 41(11 Pt 1):4432-6.

Thigpen, J.E., Haseman, J.K., Saunders, H.E., Setchell, K.D.R., Grant, M.G., and D.B. Forsythe. (2003). Dietary phytoestrogens accelerate the time of vaginal opening in immature CD-1 mice. Comp. Med. 53(6):607-615.

Thigpen, J.E., Setchell, K.D.R., Saunders, H.E., Haseman, J.K., Grant, M.G., and D.B. Forsythe. (2004). Selecting the appropriate rodent diet for endocrine disruptor research and testing studies. ILAR Journal 45(4):401-416.

Thigpen, J.E., Setchell, K.D.R., Padilla-Banks, E., Haseman, J.K., Saunders, H.E., Caviness, G.F., Kissling, G.E., Grant, M.G., and D.B. Forsythe. (2007). Variations in phytoestrogen content between different mill dates of the same diet produces significant differences in the time of vaginal opening in CD-1 mice and F344 rats but not in CD Sprague-Dawley rats. Environ Health Perspect. 115(12):1717-1726.

Vecchia, A.V. and C.G. Crawford. (2006). Simulation of daily pesticide concentrations from watershed characteristics and monthly climatic data: U.S. Geological Survey Scientific Investigations Report 2006-5181.

Vom Saal, F.S., Eve, M.D. and D.M Quadagno. (1991). Effects of maternal stress on puberty, fertility and aggressive behavior of female mice from different intrauterine positions. Physiol Behav. 49(6):1073-8.

Wang, H., Tranguch, S., Xie, H., Hanley, G. Das, S.K., and S.K. Dey. (2005). Variation in commercial rodent diets induces disparate molecular and physiological changes in the mouse uterus. Proc Natl Acad Sci. USA 102(28):9960 – 9965.

Williams, B., Higdon, D., Gattiker, J., Moore, L., McKay, M., and S. Keller-McNulty. (2006) Combining Experimental Data and Computer Simulations, With an Application to Flyer Plate Experiments. Bayesian Analysis (1):765-792.

Zehr, J.L., Van Meter, P.E. and K. Wallen. (2005). Factors regulating the timing of puberty onset in female rhesus monkeys (Macaca mulatta): role of prenatal androgens, social rank, and adolescent body weight. Biol Reprod. 72(5):1087-94.