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

OFFICE OF CHEMICAL SAFETY
AND POLLUTION PREVENTION

October 26, 2011

MEMORANDUM

SUBJECT: Transmittal of Meeting Minutes of the FIFRA Scientific Advisory Panel Meeting held July 26 - 28, 2011 on the Re-evaluation of the Human Health Effects of Atrazine: Review of Non-cancer Effects, Drinking Water Monitoring Frequency and Cancer Epidemiology

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 July 26 - 28, 2011. 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, Drinking Water Monitoring Frequency and Cancer Epidemiology.

Enclosure

cc:

Stephen Owens
Louise Wise
Vicki Dellarco
William Jordan
Margie Fehrenbach
Keith Matthews
Donald Brady
Jack Housenger
Joan Harrigan-Farrelly
Lois Rossi
Jay Ellenberger
Karen Whitby
Jack Fowle
Richard Keigwin
Anna Lowit
Elizabeth Mendez
Carol Christensen
Chester Rodriguez
Ralph Cooper
Mary Frankenberry
Nelson Thurman
Dale Kemery
Enesta Jones
Douglas Parsons
Vanessa Vu
OPP Docket

FIFRA Scientific Advisory Panel Members

Janice Chambers, Ph.D., DABT, Fellow ATS
Stephen Klaine, Ph.D.
Daniel Schlenk, Ph.D.
Kenneth Portier, Ph.D.
Susan Akana, Ph.D.
Frank Bove, Sc.D.
Richard Coupe, Ph.d
Penelope Fenner-Crisp, Ph.D., DABT
Robert Gilliom, M.S.
Ellen Gold, Ph.D.
Richard Greenwood, Ph.D
Daniel Griffith, Ph.D.
William Hayton, Ph.D.
Nelson Horseman, Ph.D.
Travis Jerde, Ph.D.
Herbert Lee, Ph.D.
James McManaman, Ph.D.
Bette Meek, Ph.D.
Kevin O'Byrne, Ph.D.
Katherine Roby, Ph.D.
Barry Timms, Ph.D.
Heather Young, Ph.D., M.P.H.

SAP Minutes No. 2011-05

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, Drinking Water Monitoring Frequency and Cancer Epidemiology

July 26 – 28, 2011
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. 2011-05

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, Drinking Water Monitoring Frequency and Cancer Epidemiology

July 26 – 28, 2011
FIFRA Scientific Advisory Panel Meeting
Held at the
Environmental Protection Agency Conference Center
Arlington, VA

Daniel Schlenk, Ph.D.
FIFRA SAP Session Chair
FIFRA Scientific Advisory Panel
Date: OCT 2 1 2011

Joseph E. Bailey
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: OCT 2 1 2011

Federal Insecticide Fungicide and Rodenticide Act
Scientific Advisory Panel Meeting
July 26 - 28, 2011

PARTICIPANTS

FIFRA SAP Chair

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

FIFRA SAP Session Chair

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

Designated Federal Official

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

FIFRA Scientific Advisory Panel Members

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

Stephen Klaine, Ph.D., Professor, Department of Biological Sciences, Director, Institute of Environmental Toxicology, Clemson University, Pendleton, SC

FIFRA Science Advisory Board Members

Susan F. Akana, Ph.D., Instructor, Department of Biological Sciences, City College of San Francisco, San Francisco, CA

Frank J. Bove, Sc.D., Senior Epidemiologist, Division of Health Studies, Agency for Toxic Substances & Disease Registry (ATSDR)/CDC, Atlanta, GA

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

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

Robert J. Gilliom, M.S., Chief, Pesticide National Synthesis, National Water Quality Assessment Program, U.S. Geological Survey, Sacramento, CA

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 Emeritus, School of Biological Sciences, University of Portsmouth, Portsmouth, United Kingdom

Daniel A. Griffith, Ph.D., Ashbel Smith Professor, School of Economic, Political and Policy Sciences, University of Texas at Dallas, Richardson, TX

William Hayton, Ph.D., Professor Emeritus, Division of Pharmaceutics, The Ohio State University, Columbus, OH

Nelson D. Horseman, Ph.D., Professor, Department of Molecular and Cellular Physiology, Program in Systems Biology and Physiology, University of Cincinnati, Cincinnati, OH

Travis Jerde, Ph.D., Assistant Professor, Indiana University School of Medicine, Department of Pharmacology and Toxicology, Indianapolis, IN

Herbert K. H. Lee, Ph.D., Professor, Department of Applied Mathematics and Statistics, Vice Provost for Academic Affairs, University of California - Santa Cruz, Santa Cruz, CA

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 Institute for Population Health Risk Assessment, University of Ottawa, Ottawa, Ontario, Canada

Kevin T. O'Byrne, Ph.D., Professor, Department of Anatomy and Human Sciences, Division of Reproduction and Endocrinology, School of Biomedical and Health Sciences, King's College London, Guy's Campus, London, United Kingdom

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

Barry G. Timms, Ph.D., Professor, Division of Basic Biomedical Sciences, University of South Dakota, Vermillion, SD

Heather A. Young, Ph.D., M.P.H., Associate Professor and Associate Chair, Department of Epidemiology and Biostatistics, The George Washington University School of Public Health & Health Sciences, Washington, DC

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, Drinking Water Monitoring Frequency and Cancer Epidemiology. Advance notice of the meeting was published in the *Federal Register* on May 6, 2011. The review was conducted in an open panel meeting held in Arlington, VA, on July 26 - 28, 2011. Dr. Daniel Schlenk 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 human health re-evaluation has involved three SAP meetings in 2010 and one in 2011. The first meeting was held in February 2010 during which the Agency presented its preliminary reviews of several atrazine epidemiology studies on birth outcomes and described a project plan to evaluate atrazine epidemiology data from the Agricultural Health Study. The second meeting held in April 2010 focused on 1) a preliminary review of experimental toxicology studies from laboratory mammals and *in vitro* studies and recent advances in understanding atrazine's mode of action along with 2) statistical and modeling approaches for evaluating monitoring frequency in community water systems (CWS). The September 2010 meeting built on the scientific analysis and SAP feedback from the April meeting. Specifically, the Agency presented scientific analyses on an empirical approach for estimating internal dosimetry and calculation of benchmark dose estimates for purposes of deriving points of departure. In addition, EPA presented a general strategy for designing a monitoring study to characterize drinking water exposures and discussed different methods for analyzing and interpreting monitoring data collected at different sampling frequencies. The September 2010 meeting also provided proposals for updating the critical duration(s) of exposure based on the new science and a preliminary evaluation of potential susceptibility of the young. In addition, the September 2010 meeting included evaluation of non-cancer epidemiology studies, a weight of the evidence evaluation of the non-cancer epidemiology studies with experimental laboratory studies and a proposal to use the non-cancer epidemiology studies qualitatively in evaluating the human relevance of experimental toxicology findings.

The July 2011 SAP meeting built on the scientific analyses and SAP feedback from the previous three SAPs by proposing a conceptual framework for the evaluation of atrazine human health non-cancer effects. This proposed conceptual framework integrated information on mode of action and adverse outcome pathways, potentially susceptible lifestages/subpopulations, drinking water exposure, internal dosimetry and water monitoring sampling uncertainty. The Agency solicited comment on the overall integrative approach proposed for atrazine along with technical considerations for each scientific component. To illustrate the conceptual framework, a case study was presented that demonstrated an approach for estimating non-cancer risk to atrazine based on an internal dose metric for temporally, spatially and demographically explicit information. The Agency also provided a preliminary review of cancer epidemiology studies, including the Agricultural Health Study, and a draft weight of the evidence (WOE) analysis on cancer that integrated mode of action, experimental toxicology information and epidemiology.

The draft WOE analysis followed the Draft Framework for Incorporating Epidemiologic and Human Incident Data in Health Risk Assessment, which was reviewed by the SAP in February, 2010. And finally, the Agency presented information regarding the continued development of the statistical analysis of drinking water monitoring data.

Opening remarks at the meeting were provided by Steven Bradbury, Ph.D., Director, Office of Pesticide Programs and John (Jack) R. Fowle III, Ph.D., Associate Director, Health Effects Division, Office of Pesticide Programs. Agency presentations were given by Elizabeth Mendez, Ph.D., Carol Christensen, Ph.D. and Chester Rodriguez, Ph.D., Health Effects Division; Nelson Thurman, M.S., Environmental Fate and Effects Division, all of the Office of Pesticide Programs; and Ralph C. Cooper, Ph.D., National Health and Environmental Effects Research Laboratory, Office of Research and Development.

PUBLIC COMMENTS

Oral Statements were presented as follows:

Janis McFarland, Ph.D., Charles Breckenridge, Ph.D., Paul Hendley, Ph.D., James Simpkins, Ph.D. (University of North Texas), Tony M. Plant, Ph.D. (University of Pittsburgh), and Harvey Clewell, Ph.D. (The Hamner Institutes for Health Sciences) all on behalf of Syngenta Crop Protection
Wendelyn Jones, Ph.D., on behalf of CropLife America
Scott Slaughter on behalf of the Center for Regulatory Effectiveness
Jere White and James C. Lamb, Ph.D., on behalf of the Triazine Network
Sarah Gallo on behalf of the National Corn Growers Association
Tyler Wegmeyer on behalf of the American Farm Bureau Federation
Stephanie Whalen on behalf of the Hawaii Agricultural Research Center

Written Statements were provided by:

Jean Public
Danielle Destefano
Lydia Lambert
Dan Campbell on behalf of Syngenta Crop Protection
Robert M. Gould, MD, on behalf of Physicians for Social Responsibility

SUMMARY OF PANEL DISCUSSION AND RECOMMENDATIONS

Charge Issue 1a - Reasonable representation of bias factor for watershed size and water body type

The Panel noted that there is no simple answer to the question posed because there are insufficient data for an adequate range of sites and years to reliably address the problem. Furthermore, as shown by available calculations of site-year bias factors, each individual site has unique characteristics that govern the adequacy of different sampling frequencies. Accordingly, the Panel offered for consideration three approaches to address the problem. The first approach is to conduct an evaluation of "homogeneous" groups to develop a categorical system of bias factors. If there are usable, discrete groups, as opposed to a continuum, then perhaps 30 sites in each group, with 10 years of high-quality data each would be a reasonable start. Reservoirs, however, which account for a large proportion of the Community Water Systems (CWS), will probably be difficult, if not impossible, to categorize because of highly variable characteristics, such as volume and residence time, which are not readily obtainable. The second approach is regression of bias vs. explanatory variables, such as basin characteristics and water-body type, thus expressing bias as a continuum governed by specific basin characteristics. This approach could be promising for at least certain parts of the problem, such as watershed size for flowing streams, but more data for multiple years will be needed at selected sites, as well as at additional sites with intermediate basin sizes. The third approach would use characteristics of a "worst case group," such as small basins, to yield a conservatively high bias factor for protective screening that would trigger monitoring. This is a practical approach that can be used now, because there is relative confidence that flowing water sites with small basins, such as the Atrazine Ecological Effects Monitoring Program (AEEMP) sites and other small-basin sites, define the worst case bias factors, both for larger flowing streams and for reservoirs, at least regarding the short duration concentrations on which the Agency is focusing. There are a significant number of CWS within the watershed size range of the AEEMP sites and, thus, the AEEMP sites are useful and applicable for defining characteristics of small-basin CWS sites in vulnerable settings. A remaining overall weakness for applying these approaches is the lack of sufficient multi-year data. Thus, it is important to continue systematic monitoring at selected sites.

Charge Issue 1b - Bias factors based on individual sites and years versus percentiles of averages across sites/years

The fundamental unit of exposure assessment, given the exposure conditions emphasized by the Agency for atrazine, is each individual site-year combination. Because each CWS site is unique, analysis generally needs to focus on each individual CWS as a unit. It is important to compute bias factors for an individual site for a particular year, and then each of these bias factors can be compared across years and across sites. Alternatively, combining data across sites and/or years before computing the bias factor will smooth out variability, downwardly biasing estimates of exposure at the peak. The condition of concern is when the maximum of a selected rolling average duration exceeds a level of concern, and this tends to be most likely in high use seasons during years when runoff after applications is high. Bias factors, to the extent they are

used for screening-level analysis, should be developed with the objective of identifying sites that merit direct monitoring. They may also be useful as a simple and transparent approach to estimating exposure for sparsely monitored sites for other purposes, such as for large-scale risk assessments or correlation with epidemiologic results. In these applications, the starting point for analysis and the endpoint of interest is the individual site, not groups of sites. However, there may be certain data analysis approaches that use data from groups of sites to make inferences for individual sites. This can be done as long as the uncertainties in predictions for individual sites are properly represented.

Charge Issue 2a - Agency's method of estimating time series using conditional simulations of variograms for monitoring data sets having 7-day sampling frequency

The Agency's method acknowledges the serial correlation latent in time series data. Conventional Box-Jenkins type ARIMA (autoregressive integrative moving average) models require uniform spacing in time, but more effectively address seasonality. Restricting attention to the days of interest appears to handle the stationarity issue in an effective way. However, sample atrazine data in Table D1.1 in Appendix D of the EPA Issue Paper imply that Julian days 101-200 may be the wrong time interval; the start time seems to be closer to Julian day 130, and seems to go beyond Julian day 200; the most recent Syngenta data support this contention for some of the other watersheds. The complication here may well be that different CWS will have different Julian day time periods (i.e., geographic variation in the windows across CWS). Any methodology that focuses on mean responses (e.g., moving averages) tends to underestimate peak atrazine concentration. Virtually all software packages report standard errors for the case of random sampling. The assumption that they are the same for systematic or stratified random sampling or for the observed non-probability sample of monitored days may well seriously impact upon the uncertainty assessment. In addition, the assumption that unsampled days are missing at random seems questionable. In contrast, the assumption that missing years for any CWS are by design, and hence eliminating those years from the population of interest, seems reasonable. Perhaps assessments within the context of mixed modeling could furnish insights here. The Panel noted other significant concerns related to possible impacts of the following issues: 1) misspecification of the semi-variogram model (a wide range of forms should be examined); 2) assuming that atrazine concentration data conform to a log-normal rather than another extreme value distribution (which can be dramatic in terms of the estimation of 95th percentiles); 3) ignoring spatial autocorrelation (which are less on mean patterns, and much more on 95th percentile estimates through effects on variance); 4) assuming a linear relationship between atrazine concentrations and "covariates" (again, potentially more dramatic on 95th percentile estimates, and adding to uncertainty rather than improving estimability); and 5) not performing a sufficient number of simulations to adequately establish the uncertainty distributions (again, especially on 95th percentile estimates rather than on mean patterns).

Charge Issue 2b - WARP (Watershed Regressions for Pesticides) modeling with longer duration sampling intervals

When monitoring data are inadequate for using kriging, time series models, or other data fill-in methods, or for reasonable sample estimates to be combined with a bias factor, WARP furnishes imputations for estimating specified concentration statistics based only on readily available watershed characteristics. WARP can be used to estimate distributional parameters, which then can be used with SEAWAVE (a parametric regression model with seasonal wave) to construct a time series that conforms, or it can be used to estimate the concentration statistic of interest (such as the predicted annual maximum 14-day rolling average) for a particular site. However, WARP is an empirical model that is subject to the same biases as the data used to build it. As such, depending on the desired application, a problem-specific version of WARP should be fit to data that are relevant to specified objectives. This perspective is consistent with previous SAP recommendations. Recognizing that EPA seeks reasonable estimates of exposure to atrazine from limited data, fine tuning of the Agency's current approach may yield a number of benefits, whereas diminishing returns in additional accuracy of atrazine estimates almost certainly will result as the complexity of the methodology increases. Furthermore, as methodological complexity increases, chances of user error also increase; the final methodology needs to be implementable by various EPA scientists with a diverse set of expertise. In other words, EPA must establish acceptable trade-offs between theory and practice in its assessment.

Charge Issue 3a - Modeling approaches for interpreting less frequently sampled monitoring datasets

When the data are sparse, it will not be possible to fit a variogram or to get accurate results solely from kriging or linear interpolation. One approach explored by EPA uses flow as a covariate. However, the initial results are not promising. There appear to be more complex relationships than a simple linear relation with flow. Precipitation might be more useful than flow as a predictor. WARP was developed to better model this situation, and it seems like it would be better to use WARP than to try to re-invent it. PRZM (Pesticide Root Zone Model) predicts chemical movement in surface soil, yielding a daily time series of potential runoff event-based concentrations, and requires more input than the WARP model. EXAMS (Exposure Analysis Modeling System) predicts the fate, transport and exposure concentration in surface water and also requires more input than the WARP model. Finally, the mass-balance model produces atrazine discharge mass quantities that often differ by orders of magnitude in either a positive or a negative direction. Consequently, WARP combined with a time-series model such as SEAWAVE, to produce a time series from the predicted statistics, appears to be a reasonable choice for obtaining supplemental data.

Output from the model yielding the best estimate of daily atrazine concentrations should be employed as the covariate in kriging. If the output for any single model appears to be no better than that for any other model, perhaps a weighted average of daily model output could be utilized. Theoretically, if no relationship exists between the model-generated data and the observed monitoring data, then the deterministic values do not impact upon the kriged values. As the relationship between the deterministic model output and the monitoring data increases in

strength, increasingly more information can be borrowed from the deterministic model output to complete each daily time series. This procedure is far superior to linear interpolation. Syngenta's approach using PRZM appears promising and the Panel encouraged that it be explored further. One could take a conservative approach using PRZM to help fill in large gaps and then account for the bias from infrequent sampling. PRZM was not initially intended for this sort of use, but it seems to perform reasonably well. SEAWAVE is an alternative model which is closer in design intention to the atrazine monitoring problem than is PRZM, and also could prove helpful. Another alternative is to consider approaches that learn from the models without needing to run the models for future years. For example, one could try regression of the sampled concentration time series on the PRZM time series, using an assumed polynomial or non-linear model or using some robust smoothing approach such as LOESS (locally weighted polynomial regression), properly lagged.

Charge Issue 4a - Use of existing atrazine monitoring data to characterize range in year-to-year variability in atrazine/TCT concentrations

Whether there are sufficient data or not depends somewhat upon the objective. If the objective is to estimate a conservatively protective bias factor to apply to sparse monitoring data, then the collection of available data, including the Atrazine Monitoring Plan (AMP) data, could be used to derive reasonable criteria that would trigger detailed monitoring of a CWS. However, if the objective is to replace detailed monitoring with predictions, the range of variability between sites and years indicates that more data would be needed to develop useful quantitative relationships, and that this would need to include detailed watershed data on a year-by-year basis. Additionally, the major drivers of atrazine concentrations in surface water have the potential to change over time and from year to year; therefore, monitoring is necessary for the foreseeable future; however, it is probably not necessary to monitor everywhere at all times. It has been demonstrated many times in chemographs presented by EPA and Syngenta how easy it would be to underestimate or completely miss a high concentration based on a small sample set. Using a model such as WARP which uses basin dependent variables and rainfall is recommended as a screening tool for inclusion in the program. One of the weaknesses in some of the historical datasets is that total chlorotriazines were not measured. Another important factor is various sources of error, whether or not they are multiplicative and how that might affect interpretation of data. There is a tremendous amount of data available, and further work should be done to characterize the site-to-site and year-to-year variability and the factors that influence the differences. Another important point is that the collection of some daily monitoring data needs to continue into the future so that current datasets are available that reflect changes in the factors that control runoff.

Charge Issue 4b - Use of PRIZM hybrid model

EPA should continue to pursue this approach because of the broad range of potential benefits of having a validated PRZM model for pesticide analysis. However, it should probably be viewed as a longer term potential tool and should not get in the way of moving ahead with the simpler transparent tools already available. An added benefit of continued development of PRZM is that the same development of watershed data, such as weather data and soils, can also

be used for site-by-site time series modeling. The hybrid model shows promise; however, PRZM is an edge-of-field model and not meant for larger basins and it does not model hydrology explicitly.

Charge Issue 4c - Other approaches to capture year-to-year variability

The Agency could try to apply the SEAWAVE time-series model which is developed using historical records of stream flow, precipitation, and temperature. Over the course of the series of SAP meetings since April 2010, there has been a shift in focus from a single maximum dose to area under the curve (AUC) for exposure durations of 4+ days; the complex approaches advocated by previous panels may not be as necessary for more frequently monitored water systems (e.g., weekly data for 4-day moving averages), and previous SAP recommendations should be considered in that context.

Charge Issue 5 - Brief exposures to low levels of atrazine and LH surge reduction and observed adverse effects

The most recent data from EPA ORD NHEERL has shown that a 4-day treatment with atrazine at 100mg/kg/day attenuated the estradiol-induced luteinizing hormone (LH) surge by 54% (AUC) in ovariectomized Sprague-Dawley (SD) rats (Goldman et al., 2011). While peak surge levels of LH were inhibited in these animals, it was marginally not statistically significant ($p=0.06$). Furthermore, 2-day treatment with this dosage of atrazine did not significantly affect the estradiol-induced LH surge (Goldman et al., 2011). The general consensus of the Panel was that the spontaneous LH surge is likely to be much more vulnerable to perturbation than gonadal steroid-induced LH surges in ovariectomized rat models, because the spontaneous surge is dependent on endogenous production of estradiol from the ovary driven by an appropriate pulsatile LH signal. Thus, it is not unreasonable to suggest that the spontaneous LH surge is highly resistant to atrazine given that 10 mg/kg for 4 days was without effect. Furthermore, it is reasonable to conclude that a 4-day exposure to 100 mg/kg is unlikely to have adverse effects on ovarian cyclicity or puberty. Syngenta's initiative to expand the animal models to the rhesus monkey to investigate the pharmacokinetics and reproductive effects of atrazine is highly commendable.

Although there was a consensus among the Panel that it is highly unlikely that the dose of atrazine under discussion (100 mg/kg for 4 days) would have adverse reproductive outcomes, it was recognized that the outcome of repeated doses, e.g., a second dose occurring 10 days later, was unknown. Concerns were also raised about the adverse effects of atrazine on body weight and food intake and the possible central nervous system (CNS) mechanisms underlying this response, which may have an impact on the reproductive systems. There was considerable disquiet among the Panel members that despite solid evidence for the mode of action (MOA) for atrazine to attenuate the LH surge, there was a complete lack of knowledge of the underlying neural or molecular mechanisms in the hypothalamus and an absence of direct coupling of LH surge attenuation to Adverse Outcome Pathways (AOP). There was a general consensus that more detailed experimentation is required, especially examining the effects of atrazine on the

CNS *per se* and the use of different routes of administration including direct application into the brain.

Discussion of the "real risk levels" of atrazine revealed a consensus among the Panel that animal experimentation (e.g., monitoring pubertal timing, spontaneous estrous cyclicity, LH surge) needs to be conducted using a dose range of atrazine that includes average and maximum levels found in water and/or levels predicted or found in humans, given chronically or intermittently. These studies should include primate models in addition to rodents. Nevertheless, a note of caution was raised by some Panel members that even if there was a lack of effect on various reproductive parameters monitored in experimental animals with 3ppb atrazine in drinking water (the current Maximum Contaminant Level), it is of some concern when considering the fact that three million people were exposed to this level and a true understanding of the AOP with these hypothalamic events is lacking.

Charge Issue 6a - Relevance of one day exposure to elicit adverse outcome(s) and significance of increase vs. decrease in LH

The primary question raised by the Agency is what may be the potential relevance of a single isolated exposure to atrazine on the LH surge and downstream LH-mediated events? A single isolated exposure to atrazine resulting in an increased LH surge is not likely to have adverse effects on downstream outcomes. If only the LH surge was modified and all other events occurring earlier in the cycle and leading to the LH surge were unaffected (including the normal process of follicle development and establishment of a population of preovulatory follicles), ovulation would likely occur as if the LH surge was 'normal' and not increased. If a single isolated exposure were to have the effect of significantly inhibiting the LH surge, downstream events would be affected including the inhibition of ovulation. Neither augmentation of the LH surge nor inhibition of the LH surge, if occurring just once, is likely to have adverse effects on fertility in women. Therefore, a single isolated exposure resulting in either increased or decreased LH surge only is not likely to negatively impact fertility. On the other hand, long term exposure where dynamics of the LH surge might be altered over an extended period of time (many months) or for an extended period of time during different stages of the reproductive lifespan (puberty, menopause) may potentially elicit adverse outcomes.

Charge Issue 7a - Human relevance of atrazine-mediated prostatitis in rats in overall hazard characterization

Several unresolved questions limit the conclusions that can be drawn regarding the human relevance of atrazine's effects in the prostate. Results in rodent models are of limited applicability due to inconsistencies in atrazine exposure levels and methodology. Secondly, the inflammation atrazine causes in rodents has not been sufficiently characterized regarding molecular and cellular events that may indicate critical changes to the tissue microenvironment, leaving open the possibility that lower doses of atrazine could produce subtle but biologically significant effects. The cellular signaling mechanisms involved have not been elucidated and molecular events such as DNA damage and imprinting changes may be possible at low levels. Such changes may accumulate during the aging process of men. Finally, it is unclear what

effects atrazine may have on the truly relevant measures of prostate cancer effects, including cancer grade, progression, and aggressiveness. The data regarding atrazine and effects on the prostate remain inconclusive, both in human and in rat models; prostatitis studies performed in rats are also inconclusive due to the irrelevance of the extremely high doses used in the study, relative to actual human exposure. While the causes of inflammation in the human prostate have not been elucidated, possibilities include viruses, systemic changes resulting from obesity, type 2-diabetes and metabolic syndrome, dietary factors and exposure to environmental factors including endocrine disruptors. Considering the role of inflammation in prostate cancer and the clear clinical relevance of prostatic inflammation to public health, the Panel recommended carefully investigating any possible role atrazine may have in the etiology of prostatic inflammation with carefully-controlled dose and model-relevant studies.

Charge Issue 8a - Evaluation of data about atrazine and mammary gland development

Based on using both approaches (ranking and morphometry) in a careful manner with Sprague Dawley (SD) rats, the NHEERL study presented in the EPA Issue Paper demonstrates that: 1) both approaches produce similar conclusions and 2) any effects of prenatal atrazine exposure on mammary gland development early in life are very subtle and not measurable by any of these techniques. The new evidence does not provide any support for an effect of atrazine on mammary gland morphology. While the Charge Question is focused on resolving differences in experimental design, data gathering, and interpretation between studies from the Fenton laboratory (Raynor, et al., 2004; 2005; Enoch, et al., 2007), and those from NHEERL (Davis, et al., 2011) and Hovey (2011), these ambiguous findings need to be considered within a larger context. The more relevant question seems to be whether development of the rodent mammary glands early in life provides an adequately robust model in which to observe subtle adverse effects of a potential environmental toxicant, such as atrazine. For a variety of reasons, it seems unlikely that mammary gland morphology, standing as a surrogate for underlying developmental processes, is adequately sensitive to fulfill this role. In conclusion, it is true that the current data "adds to the weight-of-evidence that it is unlikely that atrazine impacts mammary gland development." However, the evidence of effects in some studies, combined with the known effects of atrazine on reproductive hormones, provides an important basis for continued concern and efforts to design better studies that would determine whether these hormonal effects could contribute to poor lactation, a clearly adverse outcome, in susceptible individuals.

Charge Issue 9a - Agency's weight-of-evidence analysis and potential for differential sensitivity of the young

The Panel recommended better articulation of the weight-of-evidence analysis of the existing data as well as discussing the extent to which any uncertainties in the database might bear on the determination of the potential for life stage-related differences in sensitivity, taking into account factors such as consistency, specificity and biological plausibility. Based upon its own review and assessment of the existing database as presented in the current EPA Issue Paper and in those available for discussion at the earlier SAP meetings, the Panel concluded that there is sufficient information available to reach the conclusion that the issue of differential sensitivity

has been adequately studied. The Panel agreed with the Agency's conclusion that exposure during the earlier life stages does not appear to lead to greater sensitivity, if one accepts the premise that the data on suppression of the LH surge is appropriate for use in making the comparisons. An extensive hazard database, spanning all life stages from conception to adulthood for atrazine, indicates no unique susceptibility in the developing organism. Additionally, the proposed point of departure, based upon attenuation of the LH surge, appears to be protective against adverse reproductive/developmental outcomes such as delays in onset of puberty, disruption of ovarian cyclicity and inhibition of suckling-induced prolactin release. Assessment of weight-of-evidence for potential for differential sensitivity should include explicit consideration of consistency, dose-response concordance, specificity and biological plausibility, taking into account mode of action. Given the apparent lack of early age-related sensitivity to the neuroendocrine effects that are driving the hazard assessment, the Panel offered three options for consideration with regard to the appropriate magnitude of the FQPA safety factor.

Charge Issue 10a - Sufficiency of Agency's cancer epidemiology reviews

Overall, the Panel found EPA's review of the epidemiologic evidence to be comprehensive and systematic. EPA's literature search methods were sufficiently thorough and transparent. Similarly, EPA's method of evaluating the epidemiologic studies was, in general, sufficiently comprehensive. However, the Panel had some issues with EPA's methods of assessment. First, the Panel believed that the focus of the assessment should be on individual-level studies. Ecologic studies should be evaluated only if there are compelling reasons to do so; e.g., if there are no individual-level studies for the cancer site. Moreover, because of the limitations of ecological studies for etiological inference, negative findings in ecological studies should not be used to counter positive findings in individual-level studies. Second, the focus of the assessment should not be on whether a finding is "statistically significant," given the low statistical power of most of these studies. Third, concerning biases, it is important to provide some evidence that a specific bias is likely to be present as well as the likely magnitude and direction of the bias. Fourth, the Panel took issue with the statement on page 71 of the EPA Issue Paper that states "the weight of the evidence supports that atrazine is not likely to be carcinogenic in the human population." First of all, there is considerable uncertainty and gaps in the toxicological evidence concerning whether atrazine is a human carcinogen. Second, EPA has not done a comprehensive "weight of the evidence" assessment; instead, the toxicological evidence appears to be used to nullify any positive evidence from the epidemiologic studies. Third, the evidence across cancer sites is mixed, not uniform, with some cancer sites having no evidence for an association whereas other cancer sites having at least suggestive evidence for a causal association. The Panel did not agree that a lack of strong evidence justifies a conclusion that atrazine is not likely to be a human carcinogen. Finally, although the EPA Issue Paper appropriately summarizes the epidemiologic evidence for each cancer site, the appendix is disorganized and hard to follow because it does not evaluate the evidence for each cancer site separately. Regarding the Agricultural Health Study (AHS), the Panel found that EPA summarized, in a generally comprehensive fashion, the strengths and limitations of the study. In summary, the Panel concluded that the cancers for which there is suggestive evidence of carcinogenic potential include: ovarian cancer, non-Hodgkin's lymphoma, hairy-cell leukemia and thyroid cancer. These cancer sites require follow-up studies. In addition, cancers for which there is inadequate evidence include: prostate cancer, breast cancer, liver cancer, esophageal

cancers, and childhood cancers. These cancer sites also require follow-up studies. Cancers not likely to be caused by atrazine given the currently available data include: oral, lung, colorectal, pancreas, bladder, leukemias (except hairy-cell leukemia), multiple myeloma, melanoma, kidney, larynx, and brain/gliomas.

Charge Issue 11a - Scientific support of Agency's integrative analysis with respect to experimental toxicology and epidemiology findings; uncertainties in epidemiologic findings

The integrative analysis of the experimental and epidemiologic studies and their results that is contained in Section 3.3 of the EPA Issue Paper reflects well the state of the available scientific information regarding the relation of atrazine exposure to cancer risk. The Panel raised its concern about integration when the animal mechanistic or toxicological data are lacking, and only epidemiologic data are available (e.g., with thyroid cancer), or if the animal mechanistic or toxicological data are inconsistent or reflect different mechanisms from those in human epidemiologic data (e.g., with breast and ovarian cancer). As the Panel struggled with integrating the findings from the assessment of the human epidemiology studies, there was a realization that very little had been provided in the way of hypothesized adverse outcome pathways leading to cancer outcomes. Pathways leading to breast cancer and prostate cancer in animals are known. The human epidemiology research suggests other cancer targets, such as the thyroid, and AOPs for these cancer outcomes need to be hypothesized and related *in vitro* and *in vivo* research assessed. Many on the Panel believed that the epidemiology data failed to provide compelling evidence that atrazine is not carcinogenic. While *in vitro* and *in vivo* studies can inform epidemiology, human epidemiology findings can inform the need for additional *in vitro* or *in vivo* studies to explore hypothesized modes of action that were possibly not considered before the full epidemiology assessment was completed. In this way, the human epidemiology informs which molecular target, cellular and tissue/organ study data needs reviewing. Weak epidemiology studies with findings that suggest a strong relationship between cancer and atrazine (as is the case with the one thyroid study) should lead to additional human epidemiology research or detailed animal studies and should not be ignored because it may not fit the currently hypothesized mode of action derived from animal experiments in which the mode of action may be different than in humans. The whole point of the framework exercise is to identify gaps where lack of data may lead one to miss important health effects. A second issue that was discussed was that the EPA Issue Paper focused almost exclusively on a single mechanism of action, the neuroendocrine pathway and suppression of the LH surge, which largely has relevance only to some reproductive outcomes and a few other potential outcomes. It was recommended that a wider search for potential mechanisms for such potential cancer mechanisms as those for thyroid cancer and hairy-cell leukemia be undertaken. Again, many on the Panel believed that if the epidemiology was being used to exclude consideration of other MOAs, then this decision might not be supported by the epidemiology assessment. The Panel provided a brief summary and review of integration of results by cancer type to support their conclusions.

Charge Issue 11b - Newer epidemiology literature and justification for changing Agency's conclusions that atrazine is not likely to be carcinogenic to humans

The Panel recommended adjusting the conclusion that atrazine is unlikely to be carcinogenic to humans to "inadequate information to assess carcinogenic potential." This category is appropriate given that an association is unlikely with some cancers, but epidemiologic evidence suggests possible associations with ovarian and thyroid cancers. The Panel noted that, if possible, it would be useful and appropriate to make conclusions for individual cancers as opposed to making a blanket determination for cancer in general. Given the strong association with thyroid cancer in the epidemiologic study, a categorization of "suggestive evidence of carcinogenic potential" would be appropriate until other studies are conducted. In addition, given the suggestive association with ovarian cancer in the AHS, it would also be appropriate to use the category "suggestive evidence of carcinogenic potential."

The Panel applauded the Agency for initiating a cultural change in developing a framework to incorporate epidemiologic study results into risk assessments; however, they felt the Agency had not yet fully embraced this cultural change. The inconsistency of animal mechanistic and toxicological data with results from human epidemiologic data does not mean the risk associations identified in human studies do not reflect reality, even though animal experiments are not available or do not support the epidemiologic findings because animal models do not always apply to humans even when they are available. Notable epidemiologic findings (using the framework established in February 2010) should be given greater weight in risk assessments and should suggest avenues for future mechanistic and toxicological investigations if these are lacking, as is often the case. The Panel also noted that this seems achievable and particularly important in light of the uncertainties in existing epidemiologic evidence, to significantly expand epidemiologic information by evaluating existing information on atrazine in public water supplies, populations served and patterns in health outcomes. This approach is especially relevant for atrazine because drinking water is a primary avenue of exposure. The Agency may make significant inroads by conducting such a study in collaboration with epidemiologic researchers.

Charge Issue 12a - Strengths and weaknesses of simplified pharmacokinetic modeling approach for human extrapolation

There are advantages to using the simplest model that fulfills the purpose (in this case, prediction of toxicity of atrazine to humans on the basis of rat data), since this has the least data requirements. Where the area under the plasma concentration-time curve provides a more appropriate measure of exposure of all tissues (both target and non-target) to toxicant and metabolites than applied dose, the single compartment model (based on total radioactivity in the plasma) may be adequate. Generally, the data conform to one-compartment model behavior, and it is possible to obtain model parameter values in rat, monkey, and to a limited extent, in human. Further, the use of radiolabel facilitates the assessment of the mass balance and provides reassurance that whole dose has been accounted for. While there are advantages in using this simple model, it tends to obscure some details of the pharmacokinetics (e.g., saturable binding,

transport, and metabolism). In general, the Panel supported the work of the Agency in pursuing a dose-response analysis based on an internal dose metric, but noted that the Agency proposed the simple model as an interim approach pending evaluation of the physiologically based pharmacokinetic (PBPK) model being developed by the registrant.

Charge Issue 12b - Use of total radioactivity in pharmacokinetic analysis compared to available pharmacokinetic data for atrazine and metabolites with similar toxicological properties

Two commonly adopted approaches used in pharmacokinetic (PK) studies are PBPK modeling and PK modeling with fewer compartments; these offer different advantages and disadvantages. Complex (arguably, more realistic) PBPK models have the following disadvantages: require significant amounts of work to parameterize them; sensitivity of the analytical methods precludes their use for very low doses; and difficulty in demonstrating mass balance. Advantages of PBPK models include the following: concentrations of parent compound and individual metabolites measured using mass spectrometry represent the freely dissolved (toxicologically relevant) fraction [total radiolabel represents both the free and bound (the latter is unavailable for interaction with receptor sites or elimination in urine or bile), and where there is significant binding it will overestimate exposure of internal tissues to the toxicant]; and support interpretation of the data in terms of the physiology, and disruptions of its normal functioning, and thus are useful in mode of action studies.

There is evidence in the radiolabel studies presented by the Agency, and in a non-human primate study presented by the registrant, for the presence of two compartments from which first order elimination is taking place. One (a fast process) may be first order elimination of free material (atrazine and its metabolites, primarily diaminochlorotriazine (DACT)) from the plasma, and the second (a slow process) may represent elimination of bound material. Since the latter is a small fraction of the total material, the estimates of first order rate constants estimated by the Agency may be useful (though slightly biased). Further, (apart from those from two high doses in one study), the estimates of the elimination rate constants from different studies are reasonably consistent.

Caution is needed in the interpretation of the plasma profiles from the radiolabel studies presented by the Agency. Where few points (one per 24 h) are available, the profile will always appear smooth. However, the data based on very frequent sampling presented by the registrant provide a much more sound definition of the profile, and demonstrate sharply fluctuating plasma concentrations produced by oral gavage, but a much smoother profile by dosing in the diet. While both methods produced similar AUCs, oral gavage suppressed the luteinizing hormone (LH) surge but dietary dosing did not. Several hypotheses could explain this:

- total AUC over a critical threshold may be the relevant dose metric
- a short pulse of a very high concentration may be necessary to elicit a response
- maximum concentration is achieved later and goes on later (by 12 h) in dietary dosing compared with oral gavage.

Other dosing regimes (e.g., intravenous injection and the use of subdermal slow release formulations) may help to differentiate between these hypotheses.

Caution is also needed in interpreting the pseudo-steady state identified by the Agency in total radiolabel studies. There is no evidence for this in PBPK modeling work where the sum of atrazine and metabolites reaches a consistent maximum over the first 24 hours, and there is no measurable increase over the following days of dosing. The further increase after day four in radiolabel studies may reflect saturation of binding, and may not be toxicologically significant. Further, if a pseudo-steady state were required for suppression of the LH surge, it would be necessary to start dosing four days before the four day critical period in the estrus cycle. This is not the case, since dosing solely on the critical four days will produce the suppression.

Charge Issue 12c - Key aspects to consider concerning a PBPK model

Progress on development of the PBPK model (and its review) is noteworthy and welcome, given its potential to address interspecies differences and human variability more robustly on the basis of consideration of proportions of relevant putatively critical metabolites (avoiding generalization that there are similar fractions of active metabolites across species and avoiding reliance on default allometric scaling). The Agency is strongly encouraged to ensure that the model is sufficiently robust to meet their needs.

It is not an easy matter to evaluate or verify a complex model such as the PBPK model developed by Syngenta, and some of the key physiological aspects of the model need to be checked in order to establish creditability. These include the fitness for purpose of anatomical and physiological parameters used, including plasma-tissue partition coefficients. The model should be tested for a number of routes of exposure, and exposure metrics should be consistent with dose-response relationships with the apical (endpoint) toxicity. It should model the whole human population, and should be available on transparent modeling platforms that are widely available to the PK community. A useful framework for the evaluation/verification process is provided by the World Health Organization (WHO) guidance document that makes the following recommendations concerning the process for consideration of PBPK models in regulatory risk assessment:

- early and iterative involvement of regulatory risk assessors in model development
- access to both internal and independent expertise
- documentation by model developers in standard format for risk assessment applications and independent review

A table is presented in the Detailed Comments that summarizes the recommendations for checking the biological basis of the model, the model simulation of data, the reliability (e.g. sensitivity and uncertainty) and applicability of the model.

Charge Issue 12d - One-compartment linear model of total radioactivity and interspecies differences in pharmacokinetics

The one-compartment model adequately characterizes the PK behavior of radioactivity after administration of ^{14}C-atrazine to rat, monkey and humans. The parameter values generally differ among species within the bounds expected based upon allometric scaling principles (as used for the majority of xenobiotics) with the exception of monkey. However, there are some uncertainties associated with the total radiolabel approach because of interspecies variation in metabolic capabilities, and sites and degree of binding. While extrapolation can be affected using empirical allometric factors, it is not possible to carry out extrapolation on the basis of scientific physiological and metabolic information. It is difficult to compare the exposure time needed for effects in rat with that in humans. In part, this is due to differences in patterns of exposure: the same dose being given usually at a fixed time each day in rats, and variable intake in drinking water in humans. An advantage of the PBPK model is that it may lend itself to stochastic modeling studies of population pharmacokinetics and estimates of inter-individual variability, and this is not possible with studies based on total radiolabel.

Charge Issue 13a - Agency's rational for selecting 28-day, 14-day and 4-day exposure duration options; rationale for alternative durations of concern

Time to reach steady state and time to effect are not necessarily closely related. It could be coincidence that the 4-day exposure in rat is both the time to pseudo steady-state of radioactivity and the duration of exposure that suppresses the LH surge at a low dose. Since accumulation of the chloro-s-triazines is negligible when atrazine is dosed daily by oral gavage, the time to effect is apparently not controlled by the time required for the systemic concentrations of the triazines to reach a minimum critical level associated with the onset of effect, as the triazine exposure after the first daily dose is similar to that after the fourth dose. It is, therefore, more probable that the time to onset of effect is controlled by the pharmacodynamics; i.e., that the kinetics of events down-stream from the molecular-initiating event principally determine the onset of effect. It is not apparent what the appropriate duration of human exposure is in the context of setting the maximum level of exposure to prevent LH attenuation in humans. Without the relative rat vs. human effect kinetics, the conservative (science policy-based) approach would be to use the 4-day duration identified in the studies with rats. Information to identify an equivalent critical period in humans is inadequate.

Charge Issue 13b - Which exposure duration in humans most closely corresponds to duration found to cause adverse effects in rats

An important factor to consider is the minimum duration of LH attenuation that must occur before adverse toxicological effects ensue. Since critical information in this area is lacking, the conservative science policy-based approach is to avoid even a brief, transient suppression of LH, and the four-day exposure identified for rat would therefore seem appropriate.

Charge Issue 13c - Agency's approach to relate atrazine levels from water chemographs to predict human plasma triazine levels for proposed durations of concern. In particular, use of water AUC estimates to calculate time-weighted daily average of atrazine exposure

The approach is theoretically sound, and may benefit from the use of the Syngenta data since these are based on a much larger number of time points than any of the other studies. Those data will be more reliably extrapolated to humans because these data are less dependent on empirical transformations. The approach does provide an objective way of moving from rodent to human. The PK approach and data will be reliable, but the toxicological assumptions need to be examined in light of the recent data that have emerged. The use of water AUC to calculate a time-weighted daily average atrazine exposure is theoretically sound. This provides the time weighted average (TWA) concentration to which humans are exposed (providing that it is finished water), and hence, a reasonable assessment of exposure. An alternative to the water AUC is simply to average the measured water concentrations, but this would be inferior to using the AUC.

Charge Issue 14a - Case study: Use of 95th percentile of conditional simulations for upper bound on rolling average concentrations

The Panel believed that properly computing and interpreting the 95th percentile of the conditional simulations requires a clearer understanding and delineation of sources of error including sample size, spatial and temporal proximity of samples, the nature of a given phenomenon, model misspecification and measurement error. Conditional simulation does not embrace all of these sources of error and one weakness is that it is sensitive to the data upon which conditioning is made: a simulation replicates its conditioning values, on average. Frequently, the normal or log-normal distribution is the probability model of choice that is attached to the conditioning values; in other words, the conditioning values may be the means of a collection of normal distributions, one for each day in a time series. For the specimen atrazine data, a log-normal distributional assumption fails to adequately track the serial correlation in the data and furnishes a poorer statistical description for the monitoring data than selected alternative statistical distributions. When considering a 90% confidence interval (i.e., 5th and 95th percentiles), a balance should be maintained between claiming an excessive atrazine concentration when one does not exist, and failing to detect an excessive atrazine concentration when one does exist. For the purpose of atrazine impact analysis, this confidence interval focuses attention on the 95th percentile from a conditional simulation. Because rolling average concentrations are means, by definition they result from smoothing data so that peaks disappear (consequently, actual peak concentrations are underestimated). These averages are easier to predict specifically because they are means. Figures 27 and 29 in the EPA Issue Paper illustrate that 14- and 28-day rolling averages may be of little value for decision-making and monitoring purposes, even though they have relatively tight confidence intervals; although 4-day rolling averages are better, the crucial peak missed by them is a substantial peak. This point may be of less importance if the rolling averages represent duration of exposure. The situation may well improve with a change in the variogram model as well as a change in the probability model employed for the stochastic simulation. A variogram model that better captures autocorrelation

effects will better differentiate between the conditional and stochastic components. A probability model that allows more variability has the potential to better capture peaks. Without tracing the cumulative effects of the different sources of error from conditional simulation, perhaps a more representative approach would be to use the 97.5th percentile (i.e., a 95% confidence interval). Finally, the method illustrated for computing the 90% upper and lower confidence bounds for AUCwater are inconsistent with standard statistical definitions of these quantities and how the conditional simulations are run. Also, it was unclear in Chapter 7 of the EPA Issue Paper exactly how the 95th percentile curve from Figure 29a of the EPA Issue Paper is translated to an upper AUCwater time series; however, it appeared to represent a very conservative bound on concentrations.

DETAILED PANEL DELIBERATIONS AND RESPONSE TO CHARGE

Charge Question 1 - One approach for evaluating the performance of monitoring sampling designs is to simulate different sampling frequency strategies using robust (daily or near-daily sampling during the high use/runoff period) monitoring data. As noted in previous SAPs (April 2010 and September 2010), Heidelberg University's National Center for Water Quality Research (NCWQR) monitoring of selected watersheds in Ohio and the Atrazine Ecological Exposure Monitoring Program (AEEMP) monitoring of headwater streams in the corn belt provide the most robust datasets for atrazine. Using examples from these datasets, the USEPA presented a matrix approach for deriving a bias factor for different sampling frequencies (e.g., 4, 7, 14, 28 day intervals) for use in estimating concentrations for different exposure durations of concern (e.g., 4, 14, 28, 90 day durations).

a) Given that the factors are likely to vary based on watershed size and water body type, comment on the level of detail we would need to develop (e.g., flowing water vs. reservoir; small vs. medium vs. large watershed area). How many datasets would we need to analyze to provide a reasonable representation of a bias factor for each category?

Panel Response:

The bias factor approach is probably best viewed as an early step in the type of systematic decision process shown in Figure 22 of the EPA Issue Paper, albeit with some different methods of data analysis at some of the later steps in the process.

Application of a bias factor to exposure statistics calculated from simple linear interpolation of sparse monitoring data is a potential simple and practical approach to evaluating data from a variety of monitoring frequencies to get either unbiased or conservatively high-biased preliminary estimates of exposure metrics, depending on how the factor is derived. The approach is primarily applicable to sites with moderate frequency monitoring data, such as weekly or biweekly, so that initial biased sample estimates are more or less in statistical control. Quarterly data, for example, would be too sparse to use for short duration sample estimates.

There is no simple answer to the questions posed regarding the number of required datasets for each category of site because there are insufficient data for an adequate range of sites and years to reliably organize the problem. The reality, as shown by the available calculations of site-year bias factors, is that each individual site has unique characteristics that govern the adequacy of different sampling frequencies. Following are some observations about the problem:

- Each site year has a different concentration distribution compared with other sites that same year and with other years for the same site.
- These site-to-site and year-to-year differences in the temporal distribution and magnitudes of concentrations also apply, in varying degrees, to each specific exposure statistic of interest, such as a particular maximum rolling average (e.g., 4, 7, 14, or 28 day).

- Both the annual maximum of specific rolling averages and the temporal distribution of rolling averages, including total area under the curve (AUC) for selected intervals, may turn out to be important. The implication of this is that focusing only on a bias factor for annual maximums may not fill all the needs for future risk assessment. This reinforces the application of the bias-factor approach as a screening level analysis.
- Sparser sampling as compared with the actual "population of interest" (e.g., daily values) results in uncertainty in estimates and a tendency toward low bias for estimating high-end statistics. Bias tends to be in the direction of underestimation because short-lived high-concentration events have a higher likelihood of being missed with sparser sampling.
- Some broad differences among different types of sites are evident, such as according to basin size and reservoirs vs. streams, but there are not adequate sample sizes across the gradients of these conditions to quantify the relations with a sophisticated approach.
- Three potential approaches that could be taken to address the problem include the following:

1) Evaluation of "homogeneous" groups to develop a categorical system of bias factors. This is the approach referred to in the charge question. If there are usable, discreet groups, as opposed to a continuum, then perhaps 30 sites in each group, with 10 years of high-quality data each, would be a reasonable start. Reservoirs, however, which account for a large proportion of the Community Water Systems (CWS), will probably be difficult if not impossible to categorize because of highly variable characteristics, such as volume and residence time, which are not readily obtainable.

2) An alternate approach is regression of bias vs. explanatory variables, such as basin characteristics and water-body type, thus expressing bias as a continuum governed by specific basin characteristics. This approach could be promising for at least certain parts of the problem, such as watershed size for flowing streams, but more data for multiple years will be needed at selected sites, as well as at additional sites with intermediate basin sizes.

3) The third approach would use characteristics of a "worst case group", such as small basins, to yield a conservatively high bias factor for protective screening that would trigger monitoring. This is a practical approach that can be used now, because there is relative confidence that flowing water sites with small basins, such as the AEEMP sites and other small-basin sites, define the worst case bias factors, both for larger flowing streams and for reservoirs, at least regarding the short duration concentrations on which the Agency is focusing. There are a significant number of CWS within the watershed size range of the AEEMP sites and, thus, the AAEMP sites are useful and applicable for defining characteristics of small-basin CWS sites in vulnerable settings.

A remaining overall weakness for applying these approaches is the lack of sufficient multi-year data. Thus, it is important to continue systematic monitoring at selected sites. The sparse multi-year data is a problem for all approaches, although somewhat less for the "worst-case" group approach, because the extremes do not happen every year. In addition, commentary and responses to Charge Questions 7 and 10 suggest that atrazine damage may accumulate in a human system, perhaps in small, apparently inconsequential single increments, and hence, cumulative lifetime exposure is important to consider. This contention of impacts due to repetition implies a need to study annual time series over a lengthy time horizon. Consequently,

monitoring data for a long span of consecutive years, not just a single year or only a few recent years, needs to be described.

b) Please comment on the advantages and disadvantages of deriving bias factors based on analyses of individual sites and years compared to taking percentiles of averages across sites/years.

Panel Response:

The fundamental unit of exposure assessment, given the exposure conditions emphasized by the Agency for atrazine, is each individual site-year combination. Each CWS site has a unique watershed (e.g., hydrologic behavior, pesticide use, etc), a unique population of people served, and every year is different. Generally, analysis needs to focus on each individual CWS as a unit. It is important to compute bias factors for an individual site for a particular year, and then each of these bias factors can be compared across years and across sites. Alternatively, combining data across sites and/or years before computing the bias factor will smooth out variability, downwardly biasing estimates of exposure at the peak.

The condition of concern is when the maximum of a selected rolling average duration exceeds a level of concern, and this tends to be most likely in high use seasons during years when runoff after applications is high. Commonly, the extreme conditions happen one or more times every few to several years, as exemplified by the Honey Creek and Maumee River multi-year results submitted by Syngenta.

Bias factors, to the extent they are used for screening-level analysis, should be developed with the objective of identifying sites that merit direct monitoring. In this application, they can be biased in the conservative direction and used to identify individual sites with an unacceptable likelihood (to be defined) of exceeding a threshold, based on the available sparse sampling data. These sites would then be monitored more intensively to more accurately assess the actual condition. As part of refining the assessment with more intensive monitoring, it would be helpful to know how much variability there would be from multiple measurements taken by the same person at the same time, by different people at the same time, and comparing a manual sample to an autosampler. The uncertainty in each individual laboratory analysis of concentration should also be considered, and such statistics are generally available from all major laboratories as part of their standard quality assurance tracking. It is typical that the relative standard errors of measurement can be in the 10 to 30 percent range for atrazine, varying with concentration. Assuming that individual laboratories benchmark their performance to well-known standards, interlaboratory variability should not be a major additional problem. While such differences should be small relative to peak values, they are not likely to be negligible, and could seriously affect confidence band and percentile estimates.

Bias factors may also be useful as a simple and transparent approach to estimating exposure for sparsely monitored sites for other purposes, such as for large-scale risk assessments or correlation with epidemiologic results. In these applications, the starting point for analysis and the endpoint of interest is the individual site, not groups of sites. However, there may be certain data analysis approaches that use data from groups of sites to make inferences for

individual sites. This can be done as long as the uncertainties in predictions for individual sites are properly represented.

Charge Question 2 - The September 2010 FIFRA SAP on atrazine recommended combining a regression model such as WARP with either a deterministic model such as PRZM or a geostatistical approach in order to estimate a time series of atrazine concentrations from less frequent monitoring. In developing methods based on the SAP's recommendations, the USEPA was able to derive reasonable estimations of the time series for sampling intervals of 7-days or shorter using conditional simulations of variogram models without incorporating additional models. Although the simulations underestimated the 1-day maximums, they appeared to provide reasonable bounds for rolling average concentrations as short as 4 days.

a) Please comment on the Agency's method of estimating time series using conditional simulations of variograms for monitoring data sets such as the AMP CWS monitoring that have 7-day sampling frequencies.

Panel Response:

This methodology acknowledges the serial correlation latent in time series data (Note: Table D1.1 NCWQR 1995 Maumee River Data Set contains substantial temporal autocorrelation). Conventional Box-Jenkins type ARIMA models require uniform spacing in time, but more effectively address seasonality. (As an aside, the daily measures for 2011 (Syngenta report T001301-03) imply that, for finished water, an autoregressive moving average (ARMA) (1,1) model adequately describes these data[1]). Restricting attention to the days of interest appears to handle the stationarity issue in an effective way. However, sample atrazine data in Table D1.1 in Appendix D of the EPA Issue Paper imply that Julian days 101-200 may be the wrong time interval (the latitude range exhibited on the map of watersheds and CWS appearing in the EPA Issue Paper further supports this contention); the start time seems to be closer to Julian day 130, and seems to go beyond Julian day 200; the most recent Syngenta data support this contention for some of the other watersheds. The complication here may well be that different CWS will have different Julian day time periods (i.e., geographic variation in the windows across CWS).

Any methodology that focuses on mean responses (e.g., moving averages) tends to underestimate peak atrazine concentration. E-M (expectation-maximization) imputations are conditional expectations; in other words, they are means. The presence of autocorrelation implies that these conditional means are locally adjusted. Substituting conditional means into a time series for missing data values suppresses variance; they represent only a trend line. This is one reason for the underestimation of 1-day maximum concentrations, while obtaining reasonable estimates of rolling average concentrations[2]. This variance suppression also raises questions about assuming that standard time series developed by unadjusted kriging are representative of true daily time series.

[1] CWS-71 has a suspicious correlogram.

[2] Most of the ARMA autoregressive terms exceed 0.9, which is another reason the rolling averages tend to be successful; the MA terms tend to be about -0.4.

Virtually all software packages report standard errors for the case of random sampling. The assumption that they are the same for systematic or stratified random sampling, or for the observed non-probability sample of monitored days (diagnostics should be performed to evaluate the assumption of a pseudo-random unequal probability design, which appears to be at odds with their voluntary, truncated, and mixed water gathering nature) may well seriously impact upon uncertainty assessment. In addition, assuming that unsampled days are missing at random seems questionable. In contrast, assuming missing years for any CWS are by design, and hence eliminating those years from the population of interest, seems reasonable. Perhaps assessments within the context of mixed modeling could furnish insights here.

Other significant concerns are as follows:

1) misspecification of the semi-variogram model—a wide range of forms should be examined;
2) assuming that atrazine concentration data conform to a log-normal rather than another extreme value distribution (which can be dramatic in terms of the estimation of 95th percentiles);
3) ignoring spatial autocorrelation (which are less on mean patterns, and much more on 95th percentile estimates through effects on variance);
4) assuming a linear relationship between atrazine concentrations and "covariates" (again, potentially more dramatic on 95th percentile estimates, and adding to uncertainty rather than improving estimability); and
5) not performing a sufficient number of simulations to adequately establish the uncertainty distributions (again, especially on 95th percentile estimates rather than on mean patterns).

Issue #1 above is of particular concern. Empirical semi-variograms estimated with sample data can be extremely variable and unstable. Subsequent SAS 9.22 PROC VARIOGRAM results include standard error estimates for semi-variogram model parameters; many geostatistical software packages fail to report these values (because, for example, some use visual curve fitting). Those reported here highlight this degree of variability. One option is to exploit spatial autocorrelation by pooling data for similar watersheds when estimating semi-variograms; the subsequent discussion addresses this topic, too. Finally, relatively large nugget effects tend to overly smooth rolling averages; in the absence of any autocorrelation, the E-M solution is the sample mean.

This nugget property alludes to another issue associated with using a 1-dimensional geostatistical approach to model what is essentially a non-stationary time series: observed data are assumed to be known without error. The impact of this assumption is most evident in Figure 27a in the EPA Issue Paper where the 95th and 5th percentile curves from the conditional simulations coincide every seven days at the “known” sample points. These observed values are really estimates. For grab samples they are simply snap-shots of concentration at the time of sampling in the location that actually is being sampled. Inspecting the observed data from the NCWQR 1995 Maumee River Data Set in Appendix D, §1.1 reveals that substantial variability exists in the estimates for those dates with multiple samples. This variability appears over and

above the variability modeled by the nugget effect in a geostatistical model. This data feature should be confirmed and incorporated into the simulation exercises. Doing so will add additional variability to the simulation output, including at the sample values, making 95th percentile curves higher and 5th percentile curves lower.

Kriging produces the best linear unbiased predictors (BLUP), and is an effective way to deal with irregularly spaced data through time (treating it like a linear geographic landscape), as well as a time series with a sizeable amount of missing data (e.g., the selected subset of Table D1.1 data has 43% of its values missing). It also is substantially better than the simple linear interpolation used in some of the preliminary research (although some substitutions—e.g., vis-à-vis the PRZM—were conditional). However, the implemented methodology appears to suffer from a number of weaknesses. One drawback is the considerably restricted candidate set of semi-variogram models available in GEOEAS (i.e., exponential, Gaussian, spherical), which no longer is state-of-the-art software; semi-variogram trend lines portrayed in Figures D-3 and D-27 appear unconvincing. A misspecified model here is another source of the "nugget effect" (i.e., a non-zero intercept value arises from specification error). The autocorrelation in the selected subset sample time series is considerable, and appears to be much better described by a Bessel function [which more directly links to an ARMA(1,1) model], if not a stable function (which is similar to a Gaussian function, but with a distance exponent other than 2). These models, as well as other valid semi-variogram models, can be estimated with ArcInfo's Geostatistical Analysis module (Note: this software supported the research for Report MRID 48470008). They also can be estimated with SAS 9.22 PROC VARIOGRAM, or with modules from the R Project. These two latter software packages furnish analytical rather than visual model estimation routines. SAS quantifies uncertainty associated with semi-variogram model estimation (alluded to in the EPA Issue Paper, but without specificity, and differs from kriging prediction error). Table 1 below provides semivariogram specifications that were estimated with the log-transformed atrazine example time series data.

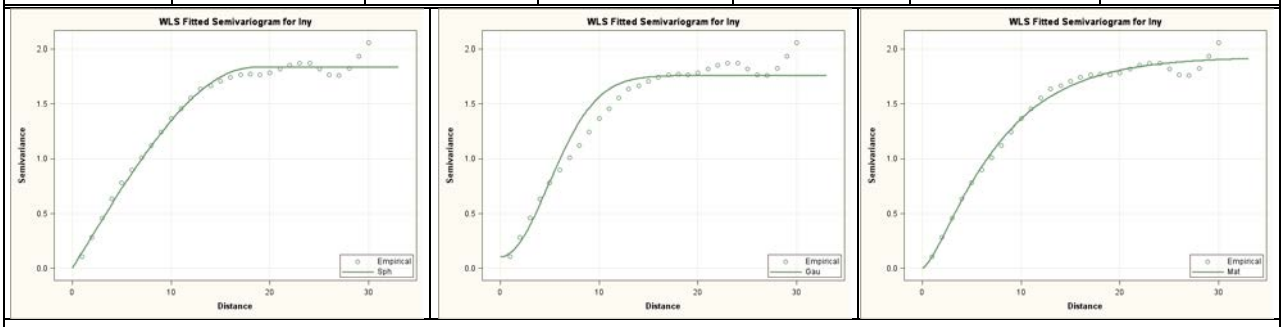
Table 1. Estimated semivariogram specifications.

Parameter	Spherical		Gaussian		Bessel	
	estimate	s.e.	estimate	s.e.	estimate	s.e.
Nugget	0.00	*	0.02	0.01	0.01	0.01
Scale	1.72	0.10	1.66	0.08	1.73	0.12
Range	21.68	1.79	8.01 (eff r = 13.9)	0.49	9.27 (eff r = 37.1)	1.54
Nu					2.80	2.56
WSSE	28.1		13.2		**12.9**	
AIC	4.0		**-18.6**		-17.3	

As shown below in Table 2, the Bessell out-performs the spherical and Gaussian models for all six additional CWS.

Table 2. Example semi-variogram model fits: CWS45 daily data. Left (a): spherical. Middle (b): Gaussian. Right (c): Bessel function.

CWS	Spherical		Gaussian		Bessel	
	WSSE	AIC	WSSE	AIC	WSSE	AIC
44	28.01	3.94	30.97	6.96	**17.36**	**- 8.41**
45	4.77	-49.16	11.59	-22.54	**1.03**	**-93.23**
52	24.81	0.30	26.35	2.11	**16.92**	**- 9.18**
54	13.23	-18.56	42.51	16.45	**11.70**	**-20.25**
71	22.78	- 2.25	37.96	13.06	**4.53**	**-48.72**
131	12.43	-20.44	25.57	1.21	**5.84**	**-41.07**



Comparable results for other semi-variogram models can be obtained with SAS PROC NLIN (utilizing weighted nonlinear least squares).

Another model misspecification concern is that estimation is being done with an overall mean when executing an "ordinary kriging" routine; this appears to be skewing results. With the atrazine time series data structure, in particular, setting the mean to zero would be better. This change in specification should solve the apparent problem with the Maumee 28-day estimation (Figure 24(d), EPA Issue Paper), as well as improve the 4-day rolling average estimation produced by simulations (Figure 26, EPA Issue Paper), eliminate the early bubbles for the daily/4/14-day averages (Figure 26), and possibly improve the confidence bands for the MO-01 data (Figure 27, EPA Issue Paper).

Nevertheless, as noted previously, kriging provides a more sophisticated approach than simple estimates from linear interpolation for sample data that have 7-day or less sampling frequencies, and appears to reduce bias, although this point was not directly addressed in the background materials. By producing a simulated complete time series, a variety of exposure measures can be assessed. However, the kriging method presented may not be the best available one for this purpose. Rather, continued consideration should be given to statistical time series models, such as the SEAWAVE model, that make use of other data as predictors, such as stream flow, precipitation, and temperature, as recommended by earlier SAPs. With a practical level of effort, these approaches can be applied. Although they may only be similar or marginally superior for 7-day and less sampling frequencies, they also can be applied to more sparse sampling data. Thus, a single approach could be applied across the continuum of CWS sampling frequencies (i.e., monitoring data availability).

Issue #2 describes a second weakness in the log-normal distribution assumption. Although the three-parameter logarithmic is the best Box-Cox power transformation enabling the selected subset of Table D1.1 data to mimic a normal frequency distribution, the transformed data still are far from bell-shaped [raw data Shapiro-Wilk (S-W) statistic = 0.898; LN(atrazine + 3.645) S-W = 0.931 (p = 0.002)]. As shown in Table 3, this same description also applies to the six recent daily sampled CWS.

Table 3. Summary diagnostic statistics for log-normal transformation of selected sample atrazine data.

CWS	Raw Data S-W Statistic	Transformation	Log-transformed Data S-W Statistic
44	0.730	LN(atrazine – 0.014)	0.962 (p = 0.011)
45	0.658	LN(atrazine – 0.019)	0.940 (p = 0.001)
52	0.710	LN(atrazine – 0.016)	0.982 (p = 0.240)
54	0.668	LN(atrazine + 0.001)	0.822
71	0.640	LN(atrazine – 0.131)	0.962 (p = 0.010)
131	0.506	LN(atrazine + 0.035)	0.816

The log-normal conceptualization describes an outcome that may be viewed as the product of many positive-valued independent random variables. It has been used to analyze extreme values of, for example, rainfall quantities and river discharge volumes, and often is acknowledged as being a heavy/fat-tailed distribution. One of the following extreme value distributions (i.e., probability distributions having extreme deviations from their medians) may well be more appropriate: Weibull, generalized extreme value, Gumbel, and Fréchet. The selected subset of Table D1.1 data better conforms to a Weibull (Anderson-Darling = 1.406) than a log-normal (Anderson-Darling = 2.059) distribution, with the largest extreme value distribution furnishing a slightly better description (Anderson-Darling = 1.356). Nevertheless, in all three cases, evidence exists suggesting that the empirical distribution differs significantly from each of these theoretical distributions. The largest extreme value goodness-of-fit appears to suffer from too many low values in the beginning of the selected subset time series. The Weibull distribution fit also suffers from anomalies in this initial part of the time series. One stated ultimate goal of the methodology is to be able to predict values greater than those sampled. The Weibull (or perhaps another extreme value) distribution offers more potential for doing this than does the log-normal distribution.

The most recent Syngenta report (T001301-03) furnishes data for an additional six CWS. These data yield the Anderson-Darling results for finished water as shown in Table 4.

Table 4. Anderson-Darling results for finished water.

CWS	First Julian day	Last Julian day	Weibull	Log-normal	n
44	91	179	1.030	3.101	89
	95	179	0.728	1.251	85
45	91	179	1.952	4.591	89
	95	179	1.440	3.984	85
52	91	179	0.998	0.584	89
54	91	179	9.533	11.142	88[a]
	132	179	2.650	2.114	47
71	91	179	0.886	1.410	89
131	91	179	8.860	10.479	89
	134	179	1.266	0.425	45

[a] a value is missing in Table 2 for F on 5/5/2011, CWS 54 because its sample bottle arrived at the laboratory broken; the data for F on 6/29/2011, CWS 71, is substituted for 6/28/2011 , a day on which the autosampler malfunctioned.

These findings support the contention that atrazine may be better described by a Weibull distribution. They also suggest that such a characterization may be watershed specific.

Issue #3 addresses the weakness of overlooking the spatial autocorrelation. This is somewhat surprising because geostatistics were developed to handle this data feature, and because of the extensive relevant discussions in Report MRID 48470008. Many geographically distributed variables within a watershed exhibit spatial autocorrelation. Time series for different watersheds also may be correlated. Depending upon such parameters as planting timing and the occurrence of storm events, watersheds of similar size with similar characteristics may well generate similar but perhaps lagged time series of atrazine concentration. If so, information can be borrowed from one time series to help complete another time series. Furthermore, information in comparable time series may be pooled to better estimate semi-variogram models.

Planned research apparently seeks to address Issue #4, namely the use of covariates (e.g., "soft" data). Co-kriging allows inclusion of additional information. One concern here is the assumption of linear relationships between atrazine and selected covariates: scatterplots appearing in Figures D-23 and D-24 in the EPA Issue Paper do not support this assumption. The furnished background reports accompanying the EPA Issue Paper state a number of R^2 values without including scatterplots: a random scatter of n-1 points of approximately the same coordinate pair accompanied by an extreme outlier can produce similar results. Some linear regression analyses involve too few points (e.g., 4, 8, 15). Further, results have been obtained with analytical routines from MSExcel; various analysts have shown many MSExcel routines to be unreliable.

Conditional simulations are an efficient and effective way to produce confidence intervals for the atrazine time series. A fifth weakness (noted as Issue #5) that pertains to these simulations is easily remedied. Simulation experiments exploit the Law of Large Numbers and the Central Limit Theorem. Those based upon 10,000 replications should be sound. Those

based upon 1,000 replications could be bolstered. Those based only upon 10 replications (e.g., Figures D-28 and D-29) are unacceptable. Except in extenuating circumstances, the number of replications should be the same across all simulation experiments.

b) Based on the USEPA's analysis using WARP with longer duration sampling intervals (Appendix D.1), what advantages does the SAP see of including WARP modeling in this approach (i.e., better estimation of the daily maximum value)?

Panel Response:

The WARP (watershed regression on pesticides) model furnishes predictions of the frequency distribution of atrazine concentrations and selected concentration statistics in specific watersheds. Its input includes the following variables: atrazine use intensity, precipitation and rainfall intensity, a soil erodibility factor, percent run off, and watershed size. Competing models include: PRZM (pesticide root zone model), EXAMS (exposure analysis modeling system), and mass-balance.

When monitoring data are inadequate for using kriging, time series models, or other data fill-in methods, or for reasonable sample estimates to be combined with a bias factor, WARP furnishes imputations for estimating specified concentration statistics based on only readily available watershed characteristics. WARP can be used to estimate distributional parameters, which can be used to construct a time series that conforms, or it can be used to estimate the concentration statistic of interest (such as the predicted annual maximum 14-day rolling average) for a particular site. However, WARP is an empirical model that is subject to the same biases as the data used to build it. As such, depending on the desired application, a problem-specific version of WARP should be fitted to data that are relevant to specified objectives. This perspective is consistent with previous SAP recommendations.

WARP model estimates of the frequency distribution of concentrations for a site can be used in combination with a time-series model, such as SEAWAVE-Q (Ryberg et al., 2010), to predict synthetic data for a 1-day temporal resolution which would allow the use of co-kriging to secure missing atrazine concentration data imputations in a time series. These supplemental data would need to be integrated properly with monitoring data. The reported experiment reveals that conditional simulations of merged WARP model-generated and sampled monitoring data are highly dependent on the WARP-based imputations. If these data are equivalent to conditional expectations, then the associated imputations will have considerably less variability (variation is suppressed), which, given the large percentage of missing values, could overweight these portions of a time series. In other words, WARP estimates alone do not really add the additional variability that would be expected in ambient measurements. Potential impacts include compromising the upper percentiles of observed concentrations, as well as reducing the likelihood of observing four-day (or x-day) rolling averages of concentration above some threshold value. Perhaps one way to temper this effect would be to add random noise to the deterministic values in such a way that they are indistinguishable from the observed monitoring data.

Uncertainty terms and variability terms need to be identified and differentiated. Time series modeling addresses structured variability issues. Uncertainty addresses issues such as misspecification of a model (i.e., not choosing the best theoretical semi-variogram function). In the next generation of the USEPA modeling efforts, thought needs to be given to directly incorporating both kinds of uncertainty in the model specification for atrazine exposure.

One goal of the time series analyses is to establish an upper percentile threshold that is not excessively conservative (orders of magnitude beyond the observed data); the final percentile should furnish adequate protection, but not far more protection than is necessary, which could deprive society of the benefits of atrazine while really not significantly improving the likelihood of avoiding adverse health effects. Within the context of this goal, error propagation merits evaluation to see whether or not compounding occurs, with this evaluation being balanced against returns on an investment of resources in such a study. Conceptual arguments in terms of plausibility may be sufficient to dismiss some propagation possibilities. Sources of error meriting consideration range from merging spatially gridded field data that are 4-by-4 km (for rainfall) and 85-by-74 km (for temperature), to raster images of rainfall totals from historical radar weather data—all of which may involve raster-to-polygon conversions—to the numerous assumptions employed by model-based imputators (e.g., a 1:1 relationship between relative percentiles of flow and atrazine for the WARP model). One concern expressed in the EPA Issue Paper is the need for a priori knowledge about reasonable upper limits for peak concentration estimates. Although such figures furnish checks for synthetic results, percentages of these peaks are not being estimated. Furthermore, because imputations are conditional means, estimation of extremes is unlikely. Replacing a log-normal probability model with an extreme value probability model may help remedy this situation. Research establishing a valid auto-Weibull type of model might be useful. For example, the autocorrelation trend in the specimen atrazine data suggests a Weibull distribution with a shape parameter of roughly 3.2 (which implies that it approximately mimics a bell-shaped curve; this may be a primary reason that the log-normal distribution performs well). With an accompanying scale parameter of 2.3 and a suitable autocorrelation factor (e.g., estimated from the observed data), the resulting daily time series resembles the observed atrazine time series: the averages of 10,000 replications of 100 draws from a Weibull distribution almost perfectly replicate the spatial autocorrelation pattern ($R^2 > 0.999$); and, the approximate 95% confidence intervals based on 10,000 replications of 100 draws gives an upper bound of roughly 23.5 for the maximum observed concentration value of 14.058. This latter result supports the need for a priori knowledge of reasonable upper limits for peak concentration estimates, as well as indicates that imputations based upon a deterministic model, or empirical regression model, such as the WARP model combined with an extreme value distribution, such as the Weibull, could allow prediction of values much greater than those sampled. This example also illustrates that imputed values tend to be highly dependent upon the deterministic model predictions (e.g., the synthetic temporal autocorrelation component employed here accounts for 87% of the variance in the observed time series, allowing much less variability for the stochastic component).

Recognizing that EPA seeks reasonable estimates of exposure to atrazine from limited data, fine tuning of the Agency's current approach may yield a number of benefits, whereas diminishing returns in additional accuracy of atrazine estimates almost certainly will result as the complexity of the methodology increases. Furthermore, as methodological complexity increases,

chances of user error also increase; the final methodology needs to be implementable by various EPA scientists with a diverse set of expertise. In other words, EPA must establish acceptable trade-offs between theory and practice in its assessment.

Charge Question 3 - Given that most monitoring data for pesticides are based on less frequent sampling intervals (e.g., bi-weekly, monthly), the USEPA also explored additional approaches that combined variogram models with covariate approaches – in particular, correlation with flow within a narrow window – or with WARP. We also considered methods of filling in time series concentrations using a deterministic model such as PRZM or a mass balance model, although such approaches would be more resource/data intensive.

a) Please comment on these additional modeling approaches for interpreting sparse (less frequently sampled) monitoring datasets.

Panel Response

When the data are sparse, it will not be possible to fit a variogram or to get accurate results solely from kriging or linear interpolation. One approach explored by EPA uses flow as a covariate. However, the initial results are not promising. There appear to be more complex relationships than a simple linear relation with flow. A scatterplot of flow versus atrazine concentration shows two distinct branches, one of which involves a correlation between flow and atrazine concentration, the other of which shows no relationship (i.e., when rainfall occurs prior to field application, or after most of the pesticide has already been washed away by previous rainfall events). Precipitation might be more useful than flow as a predictor. WARP was developed to better model this situation, and it seems like it would be better to use WARP than to try to re-invent it.

As noted previously, the WARP model furnishes predictions of the distribution of atrazine concentrations in specific watersheds. Its input includes the following variables: atrazine use intensity, precipitation and rainfall intensity, a soil erodibility factor, percent run off, and watershed size. PRZM predicts chemical movement in surface soil, yielding a daily time series of potential runoff event-based concentrations, and requires more input (e.g., temperature, land use, soil) than the WARP model. It uses spatially specific NEXRAD (next-generation radar) data, requiring additional data merging. EXAMS predicts the fate, transport and exposure concentration in surface water by combining chemical loadings, transport, and transformation into a set of differential equations using the law of conservation of mass as an accounting principle. Its data inputs include fundamental chemical properties of atrazine, and up to 32 different segments for a given watershed. Up to 28 different substances may be simulated. The EXAMS model also requires more input than the WARP model. Finally, the mass-balance model, which describes variations in atrazine concentration as a series of storm-event associated peaks that taper off over time, produces atrazine discharge mass quantities that often differ by orders of magnitude in either a positive or a negative direction. Consequently, WARP, combined with a time-series model such as SEAWAVE (to produce a time series from the predicted statistics), appears to be a reasonable choice for obtaining supplemental data.

Output from the model yielding the best estimate of daily atrazine concentrations should be employed as the covariate in kriging. If the output for any single model appears to be no better than that for any other model, perhaps a weighted average of daily model output could be utilized. Reconsidering the semi-variogram models for the specimen atrazine data, including the synthetic spatial autocorrelation factor as a covariate for co-kriging produces considerable smoothing of the daily experimental variogram. In addition, the resulting goodness-of-fit diagnostics improve for all candidate model specifications, and furnish additional evidence that the Bessel function may be the preferred model. These results corroborate that imputed values will tend to be highly dependent upon deterministic model predictions used as covariates. The lack of a stochastic component for the imputations will tend to suppress variability (deterministic model predictions are similar to conditional expectations).

Theoretically, if no relationship exists between the model-generated data and the observed monitoring data, then the deterministic values do not impact upon the kriged values. As the relationship between the deterministic model output and the monitoring data increases in strength, increasingly more information can be borrowed from the deterministic model output to complete each daily time series. This procedure is far superior to linear interpolation.

Syngenta's approach using PRZM appears promising and could be explored further. One could take a conservative approach using PRZM to help fill in large gaps and then account for the bias from infrequent sampling. PRZM was not initially intended for this sort of use, but it seems to perform reasonably well. SEAWAVE is an alternative model that is closer in design intention to the atrazine monitoring problem than is PRZM, and also could prove helpful.

One principal weakness in using deterministic model output arises from the assumptions involved. Because values between observed monitoring points in time are unknown, they may not coincide with model output, even if the observed data perfectly align with their corresponding subset of model output. This weakness furnishes a strong argument to employ a time-interval stratified random sampling design, rather than a systematic design (with a random start weekday). Sensitivity analyses, especially with regard to error propagation, could shed light on the magnitude of impacts of certain assumptions. Critical assumptions include the following:

1. (PRZM) 60% of atrazine is applied at four uniformly distributed major pulses, with the remaining 40% being applied uniformly across all other days.
2. Movement through a watershed is indexed to that shortest path having the largest shortest path distance separating any of its outlet and its headwater pairs.
3. Growing degree days are defined by (Tmax – Tmin)/2 – 50.
4. All corn and sorghum crop areas are treated.
5. (PRZM) all watershed farmers use atrazine in a similar way to those in the baseline CRC survey.
6. (PRZM) the atrazine use rate is uniform across all soil types.
7. Watersheds experience no conservation practices, and have good hydrologic conditions.
8. Non-random sample of monitoring data can be treated like a random sample.

Assumptions that most likely have little adverse effect on results include the following:

1. Results for an irregularly spaced time series that can be adjusted by weighting each value by 50% of the time distance between it and its preceding as well as its next sample value.
2. Multiple sample values for a day that can be represented by their geometric mean.
3. Values substituted for those quantities less than the detection limit.
4. Designing analyses in such a way that concentration estimates tend to be conservative.
5. Half-life of atrazine is 61 days in a watershed (although evidence exists suggesting that it has a much longer half-life in subsurface soils).

In the end, a meaningful model is only as good as the ability of its assumptions to mirror the real world.

One notable trade-off is between the expenditure of resources to collect reliable sample data, in a design-based context, and the use of model-based techniques that require resources to assemble massive amounts of ancillary data properly and then convert them into sample data equivalences. Physical sample collection requires retrieval followed by storage of specimens, and is plagued by instrument malfunctions as well as human error. Model generated results suffer from data availability as well as human error. One cannot go back in time to correct the former; updates followed by model re-executions allow corrections to be made for the latter.

Another alternative is to consider approaches that learn from the models without needing to run the models for future years. For example, one could try regression of the sampled concentration time series on the PRZM time series, using an assumed polynomial or non-linear model or using some robust smoothing approach such as LOESS (locally weighted polynomial regression), properly lagged. If one finds that the resulting R-Square is close to 1, it would suggest that the PRZM model would be a good predictor of the concentration time series for the un-measured time points. This approach might work quite well for what Syngenta referred to as the small AMP CWS size class because the smaller areas allow for only one or a few fields to be impacting the water concentration and time lags of the order of a day or days would be reasonable. For larger CWS size classes, incorporating many more fields and longer transit times, it is more likely that the concentration time series is some weighted sum of the variably lagged WARP-based SEAWAVE model predictions for all fields impacted by significant rainfall events. This would be a very challenging model to fit. Perhaps a “sum of correlated series model” (a limit theorem model) might fit these data. This kind of regressed time series approach does not preclude the introduction of further autocorrelation structure in the residual of the model fit.

To some extent, Syngenta is taking this approach by using the modified PRZM model to “fill in the extreme events” between the sampled concentration values. One Panel member suggested that if you are going to fill in the extreme events, why not go the whole way and use WARP-based SEAWAVE predictions of the time series to fill in the rest of the sequence. This does not appear to be what Syngenta is doing, as slide 20 of the Syngenta Occurrence in Drinking Water presentation in the meeting docket (Atrazine Occurrence in Drinking Water – Monitoring Frequency and Modeling, Hendley, P. et. al, 2011) shows its scientists are using linear interpolation between the WARP-inspired maximum and the observed sample point. One

other issue with the Syngenta approach is that it assumes that an atrazine runoff event occurs between every pair of sampling days. For seven day sampling, this essentially assumes a runoff event every week, a highly unlikely occurrence given the data presented previously before this Panel (SAP, April 2010).

Ideally, a regime needs to be set up such that sites can move between different frequencies of monitoring (moving both up and down in frequency). Thus one needs an approach for upgrading sparse datasets, as well as an approach for downgrading compliant highly-monitored datasets.

In summary, the Panel encouraged further development of the PRZM approach and the creation of a database to drive this nationally. Development and data collection would have transfer potential for other chemicals, which could lead to a good return on investment for EPA. The Panel also cautioned against holding up a decision on atrazine monitoring while trying to create this larger framework and encouraged EPA to move forward with atrazine monitoring in parallel with framing the wider problem.

Charge Question 4 - In characterizing overall uncertainty in exposure estimates from monitoring data, the USEPA must also consider how many years of monitoring are necessary to provide a reasonable bound on the year-to-year variability or, alternatively, develop methods of placing the existing monitoring data in context of likely exposures that may occur over a time period of interest (for instance, 15 year cycles for registration review).

a) Please comment on the sufficiency of existing atrazine/triazine monitoring data available to the Agency – in particular the Atrazine Monitoring Program (AMP) coupled with the earlier Voluntary Monitoring Program (VMP), which conceivably span from 1993 to the present for some CWS – for use in characterizing the likely range in year-to-year variability in atrazine/TCT concentrations.

Panel Response:

Atrazine is an important herbicide used in U.S. agriculture because it is relatively inexpensive and controls a broad spectrum of weeds especially in corn and sorghum. There have been frequent detections of atrazine in surface waters of the U.S., particularly in the Midwest Corn Belt region. Atrazine also has been detected in the finished drinking water of some Community Water Systems. Its occurrence in surface waters of a basin is mostly dependent upon hydrology and application rates within that basin, and the year-to-year variability in detected surface water concentrations is related to the amount and timing of rainfall in relation to atrazine application.

Whether or not sufficient data exist about atrazine contamination depends somewhat upon the objective. If the objective is to estimate a conservatively protective bias factor to apply to sparse monitoring data, then the collection of available data, including AMP, could be used to derive reasonable criteria that would trigger detailed monitoring of a CWS. However, if the objective is to replace detailed monitoring with predictions, the range of variability between sites

and years indicates that more data would be needed to develop useful quantitative relationships, and that this would need to include detailed watershed data on a year-by-year basis.

Additionally, the major drivers of atrazine concentrations in surface water have the potential to change over time and from year to year. Climate change predictions show rainfall patterns changing, and market forces such as the Biofuels Initiative have changed the amount of corn planted, changing the amount of atrazine applied. Monitoring is necessary for the foreseeable future; however, it is probably not necessary to monitor everywhere at all times. A bias factor could be used to set criteria for more or less intensive monitoring, or other criteria could be used, such as no X day rolling average is greater than Y. Additionally, other CWS not now included in the AMP should be evaluated for inclusion. Currently, inclusion in the program is determined upon documentation of an exceedance of the drinking water standard in samples collected as a requirement for compliance with the Safe Drinking Water Act (SDWA), which is based upon 4 samples per year (most not in the high use season). It has been demonstrated many times in chemographs presented by EPA and Syngenta how easy it would be to underestimate or completely miss a high concentration based on a small sample set. Using a model such as WARP, which employs basin dependent variables and rainfall, is recommended as an additional screening tool for inclusion in the program.

One of the weaknesses in some of the historical datasets on atrazine is that total chloro-triazines were not measured. This will systematically underestimate the total concentration of atrazine derived compounds. In addition, simazine use and its detected concentrations have been increasing in some parts of the Corn Belt. Simazine sampling data need to be included for accurate assessment of potential triazine exposure and risk.

One important factor not included in this analysis is error and how that might affect interpretation of data. There will be inherent error in sampling (i.e., autosampler versus grab sample versus a depth- and width- integrated sample), perhaps changes in laboratory analytical procedures for assaying atrazine over time, and differences between laboratories. The assumption that a water sample specific to a point in time and space represents the concentration for some length of time and volume of water in a surface water body can also introduce error. Are these sources of error additive, multiplicative, or are some subsumed in others? Determining the degree of compounding of these errors may impact the size of confidence intervals.

However, that being said, there is a tremendous amount of data available, and further work should be done to characterize the site-to-site and year-to-year variability and the factors that influence the differences. Understanding year-to-year variability is difficult, because, in some years, the variability will be much higher than in others. But the question can be broken down into two components: shape and magnitude. The magnitude of the peak of the chemograph will have large year-to-year variability. However, the shape for a particular CWS may be similar across years. The number of peaks, the broadness of the peaks, and the correlation structure of the time series may not show as much annual variability.

Another important point is that the collection of some daily monitoring data needs to continue into the future so that current datasets are available that reflect changes in the factors

that control runoff. Syngenta is currently monitoring a number of CWS, and that will be a great addition to the data available for future analysis.

b) Please comment on the Agency's suggestion for using a PRZM hybrid model, calibrated on the current years of monitoring, to provide estimates for a wider time frame by modeling additional years using weather data that span a 30- to 50-year period.

Panel Response:

EPA should continue to pursue this approach because of the broad range of potential benefits of having a validated PRZM model for pesticide analysis. It should probably be viewed as a longer term potential tool, and should not get in the way of moving ahead with the simpler transparent tools already available. An added benefit of continued development of PRZM is that the same development of watershed data, such as weather data and soils, can also be used for site-by-site time series modeling. Additional comments on the use of PRZM can be found in the response to Charge Question 3a.

Syngenta has shown some innovative approaches to modeling atrazine runoff. The hybrid model shows promise. However, it should be noted that PRZM is an edge-of-field model and not meant for larger basins. PRZM does not model hydrology explicitly. This can be of importance with larger basins that receive differing amounts of rainfall over their geographic extents which might affect the timing of the arrival of atrazine at a sampling point. Also, one cannot model how changes to hydrology might affect the movement of atrazine and it is not clear how important watershed characteristics such as reservoirs can be handled. The response to Charge Question 3a also gives a list of assumptions used in PRZM that are important to the model accuracy and should be considered when evaluating it.

c) What other possible approaches can the SAP recommend for capturing year-to-year variability?

Panel Response:

The Agency should try to apply the SEAWAVE time-series model, which is developed using historical records of stream flow, precipitation, and temperature, as was recommended by previous SAPs.

It is worth noting that some of the previous SAP recommendations were made because of perceived interest in daily maximum concentrations. For the April 2010 SAP, the Panel was explicitly charged with estimating one-day peaks from weekly data. In September 2010, there was interest in the ability to estimate a variety of durations of exposures However, at the end of that meeting, there was a shift toward interest in area under the curve (AUC) for exposure durations of four or more days. With this shift from a single maximum dose to AUC, the complex approaches advocated by previous panels may not be as necessary for more frequently monitored water systems (e.g., weekly data for 4-day moving averages), and previous SAP recommendations should be considered in that context.

Charge Question 5 - In 2000, the SAP agreed with the Agency's proposal for atrazine's neuroendocrine MOA, and they further concluded that it is unlikely that the mode of action by which atrazine induces mammary tumors in adult female Sprague Dawley rats could be operational in humans. The SAP further concluded that it is not unreasonable to expect, however, that atrazine might cause adverse effects on hypothalamic-pituitary-gonadal (HPG) function in humans if exposures were sufficiently high and that perturbation of the HPG axis was relevant for developmental and reproductive effects (FIFRA SAP, 2000). In the 2003 IRED, the Agency identified perturbations of the neuroendocrine system (particularly LH regulation) leading to reproductive toxicity as the most biologically plausible and sensitive effects attributable to atrazine exposure. The adverse outcomes occurring as a consequence of disruptions to the HPG axis in rats include disruption of estrous cyclicity and delays in puberty onset (males and females). An additional effect – not directly linked to LH disruption – is the decreased suckling-induced prolactin release in milk early in life (perinatally), which leads to increased incidence of prostatitis in young adult rats. All of these effects can be linked to and/or occur at higher doses than the atrazine-induced changes in LH secretion. Research conducted since the 2003 IRED continues to point to LH surge attenuation as the most sensitive effect in the atrazine database. Consequently, the Agency will continue to use changes in LH secretion as a sentinel endpoint for HPG perturbations and the basis of the atrazine risk assessment. The September 2010 SAP " 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."

a) Currently available data show that in the rat a brief exposure (as brief as 4 days) to low levels of atrazine can elicit decreases in LH. Please comment on the biological plausibility of these brief changes leading to an adverse outcome taking into account typical variability and how long and how much a LH surge reduction is needed to cause the observed adverse effects (i.e., disruptions in cyclicity, delayed puberty, and prostatitis).

Panel Response:

The most recent data from EPA ORD NHEERL has shown that a 4-day treatment with atrazine at 100mg/kg/day attenuated the estradiol-induced LH surge by 54% (AUC) in ovariectomized Sprague-Dawley (SD) rats (Goldman et al., 2011). While peak surge levels of LH were inhibited in these animals, it was marginally not statistically significant (p=0.06). Furthermore, 2-day treatment with this dosage of atrazine did not significantly affect the estradiol-induced LH surge (Goldman et al., 2011). The question arises as to whether the LH surge reduction demonstrated by Goldman et al. (2011) could cause disruption to ovarian cyclicity and delayed puberty in the rat. It has already been demonstrated that 50 mg/kg/day (or more) atrazine starting on post natal day (PND) 22 can delay puberty determined by vaginal opening (Laws et al., 2000). However, 12.5 or 25 mg/kg/day administered from PND 22 do not affect pubertal timing or estrous cyclicity; i.e., the no observed adverse effect level (NOAEL) is

25 mg/kg/day (Laws et al., 2000). Although there are no data for a 4-day treatment with atrazine at 100mg/kg/day which examine pubertal outcome or estrous cyclicity, the fact that pubertal timing and estrous cyclicity were normal after almost two weeks of 25 mg/kg/day (starting on PND 22) atrazine (Laws et al., 2000) might suggest an unlikely effect of atrazine at 100 mg/kg/day for just four days on these reproductive parameters. Moreover, there is a huge redundancy extant in the reproductive neuroendocrine axis. It was demonstrated in the early 1970s that there is enormous variation in the peak amplitude of the spontaneous LH surge in the rat; ranging from approximately 200 - 2,000 ng/ml (Everett et al., 1973). Furthermore, over this range, the rate of ovulation, i.e., number of ova released, is not different. With respect to hypothalamic determinants of the LH surge, Bowen et al. (1998) has shown that only 10% of the gonadotropin-releasing hormone (GnRH) surge is required for a normal LH surge in the ewe. Collectively, these data might suggest that a 54% reduction in AUC (with no change in peak LH surge levels) of the LH surge in response to a 4-day treatment with 100 mg/kg/day atrazine is unlikely to affect puberty or ovarian cyclicity in the rat. In addition, Herbison et al. (2008) has shown that transgenic mice with selective reduction in the number of GnRH neurons require only 12% of the GnRH neuronal population for normal pubertal onset and first estrous cycle. The inference is that with 12% of GnRH neurons, there is a normal pulsatile secretion of LH to drive puberty and first ovulation, thus demonstrating a huge redundancy within the hypothalamic GnRH neuronal network. It is important to be cognizant that the toxicological doses of atrazine being discussed (e.g., 12.5 – 100 mg/kg/day) are not relevant to probable exposure levels in the "real" environment.

Syngenta data presented on July 27, 2011 clearly show that atrazine at 50 mg/kg for 4 days resulted in only a 50% reduction in peak amplitude of the spontaneous LH surge in SD rats. Moreover, 10 mg/kg atrazine for 4 days was without effect (i.e., a NOAEL). The general consensus of the Panel was that the spontaneous LH surge is likely to be much more vulnerable to perturbation than gonadal steroid-induced LH surges in ovariectomized rat models, because the spontaneous surge is dependent on endogenous production of estradiol from the ovary driven by an appropriate pulsatile LH signal. Thus, it is not unreasonable to suggest that the spontaneous LH surge is highly resistant to atrazine. This is consistent with the observation that treatment with 10 mg/kg for 4 days was without effect. Furthermore, it is reasonable to conclude that a 4-day exposure to 100 mg/kg is unlikely to have adverse effects on ovarian cyclicity or puberty. Syngenta's initiative to expand the animal models to the rhesus monkey to investigate the pharmacokinetics and reproductive effects of atrazine is highly commendable.

Although there was a consensus among the Panel that it is highly unlikely that the dose of atrazine under discussion (100 mg/kg for 4 days) would have adverse reproductive outcomes, it was recognized that the outcome of repeated doses, e.g., a second dose occurring 10 days later, was unknown.

Concerns were also raised about the adverse effects of atrazine on body weight and food intake and the possible central nervous system (CNS) mechanisms underlying this response which may have an impact on the reproductive system. There was considerable disquiet among the Panel members that despite solid evidence for the MOA for atrazine being the attenuation of the LH surge, there was a complete lack of knowledge of the underlying neural or molecular mechanisms in the hypothalamus and an absence of direct coupling of LH surge attenuation to

Adverse Outcome Pathways (AOP). Indeed, it was argued that its relevance to humans, or even rats, is questionable as is its use as a sentinel event for atrazine risk assessment. There was a general consensus that more detailed experimentation is required, especially examining the effects of atrazine on the CNS per se and the use of different routes of administration including direct application into the brain. However, poor solubility of atrazine may be a confounding issue. Data on the effects of atrazine on CNS function are extremely limited. Foradori et al. (2009a) have shown loss of c-fos expression (a marker of neuronal activation) in GnRH neurons with 100 mg/kg for 4 days concomitant with blockade of the LH surge. Syngenta noted that, in addition to atrazine reducing c-fos expression in the hypothalamus, there was evidence for c-fos activation in the amygdala, which is part of the limbic brain. The significance of the latter is unknown but might suggest a behavioral outcome.

Despite these concerns and given that the neural mechanisms underlying LH surge generation in the rat are different from primates, including humans (as presented by Syngenta), it was posited by a Panel member that perturbation of the LH surge per se would impact significantly on reproductive function because of its central importance in ovarian cyclicity. However, as discussed above, attenuation (approximately 50%) of the LH surge may have little impact on reproductive function with only near or complete blockade of the LH surge being detrimental, but this requires dosages of atrazine more like 100 mg/kg, which is far in excess of any reasonable dose likely to be encountered in the natural environment. In addition, experiments involving the monitoring of pulsatile LH secretion, which directly reflects hypothalamic GnRH pulse generator activity (the central regulator of the reproductive neuroendocrine axis) and critical for control of follicular development and genesis of the preovulatory LH surge, have revealed no effect with 4 days of treatment with 100 mg/kg atrazine and only a modest, though significant, reduction in LH pulse frequency with 200 mg/kg atrazine for 4 days (Foradori et al., 2009b). Discussion of the "real risk levels" of atrazine revealed a consensus among the Panel that animal experimentation (e.g., monitoring pubertal timing, spontaneous estrous cyclicity, LH surge) needs to be conducted using a dose range of atrazine that includes average and maximum levels found in water and/or levels predicted or found in humans, given chronically or intermittently. These studies should include primate models in addition to rodents. Nevertheless, a note of caution was raised by some Panel members that even if there was a lack of effect on various reproductive parameters monitored in experimental animals with 3ppb atrazine in drinking water (the current MCL), it is of some concern when considering the fact that three million people were exposed to this level and a true understanding of the AOP with these hypothalamic events is lacking.

Charge Question 6 - During the September 2010 SAP, the Panel raised the issue of single vs. multiple atrazine exposure effects on the LH surge. In their report to the Agency the SAP commented:

"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."

In response to the Panel's comments, EPA scientists in the Office of Research and Development have undertaken a series of experiments to try to elucidate the nexus between phase of the cycle and duration of exposure. This research is in the early stages. Initial results suggest that a single high dose of atrazine (100 mg/kg bw) can affect the LH surge. However, the effect seen was an increase in LH rather than the decrease observed after 4 days of exposure.

a) Please comment on the potential relevance of one day exposure to elicit an adverse outcome(s) and the significance of an increase vs. a decrease in LH.

Panel Response:

Significant data have been accumulated that demonstrate that exposure to atrazine during the rat estrous cycle for four days beginning on estrus results in the inhibition of the LH surge. Outcomes of diminution of the LH surge include inhibition of ovulation. Additional data demonstrate that during four days' exposure, systemic levels of LH pulse with each administration/exposure attain a pseudo-steady state. In order to understand the relevance of changes in LH to outcomes, it is first necessary to understand how paradigms of exposure impact the dynamics of the LH surge. For example, would a single isolated exposure to atrazine result in an altered LH surge? Further, what is the relationship between exposure on specific days of the estrous cycle and impact on the LH surge? For example, would a single isolated exposure near the onset of the LH surge have the same impact as a single isolated exposure two days prior to the LH surge? To directly address the potential relevance of a single day exposure eliciting adverse outcomes specifically monitored by changes in parameters of the LH surge, the Agency has initiated a series of studies of which preliminary results were presented to the Panel.

Studies were initiated using the ovariectomized estrogen-replaced rat model. This model system establishes an endocrine state in which an LH surge occurs in a predictable manner once daily. A single isolated exposure to atrazine just prior to the initiation of the LH surge resulted in potentiation of the surge. This is in contrast to longer exposure where the LH surge is diminished. The complexity of the endocrine system is illustrated by this result and further demonstrates the effects of atrazine on the hypothalamic-pituitary-adrenal axis and the hypothalamic-pituitary-gonadal axis and the interplay between the two systems. As demonstrated previously, exposure to atrazine results in the adrenal release of corticosterone and progesterone. A biphasic role for progesterone in influencing the LH surge has been clearly established whereby progesterone initially enhances the effects of estradiol in the induction of the LH surge and continued or extended exposure to progesterone is inhibitory to the LH surge.

The primary question raised by the Agency is what may be the potential relevance of a single isolated exposure to atrazine on the LH surge and downstream LH-mediated events? A

single isolated exposure to atrazine resulting in an increased LH surge is not likely to have adverse effects on downstream outcomes. If only the LH surge was modified and all other events occurring earlier in the cycle and leading to the LH surge were unaffected (including the normal process of follicle development and establishment of a population of preovulatory follicles), ovulation would likely occur as if the LH surge was 'normal' and not increased. If a single isolated exposure were to have the effect of significantly inhibiting the LH surge, downstream events would be affected including the inhibition of ovulation. There exists a large window-of-safety with regard to the LH surge such that normal ovulation will occur over a wide range of LH surge levels (peak values and/or AUC).

Neither augmentation nor inhibition of the LH surge, if occurring only once, is likely to have adverse effects on fertility in women. The potentially more significant of the two situations would be the inhibition of ovulation resulting in a non-fertile cycle. The length of an anovulatory cycle may be altered by a day or two; however, the following cycle would likely be normal. Over the reproductive lifespan of a woman, there are multiple situations that may result in anovulatory cycles (for example, stress). In addition, anovulatory cycles occur during the normal transitions of menarche and menopause in women. Therefore, a single isolated exposure resulting in either increased or decreased LH surge only is not likely to negatively impact fertility.

Additional questions raised during the September 2010 SAP have yet to be addressed. As the Agency was reporting preliminary data related to a single isolated exposure, these additional questions may be addressed in the larger study.

It is of interest to understand the significance of when during the cycle exposure occurs as it relates to the LH surge. This is a somewhat complex question to address in a rodent model where the estrous cycle is short; however, the question has relevance in translating the experimental data to the potential for risk to women. As described in the Goldman et al. (2011) internal report, potential significance was addressed when atrazine was administered near the onset of the LH surge. If atrazine alters the LH surge only when exposure occurs at the onset of the surge, the potential that a single isolated exposure in women would have any effect on the LH surge is low.

The results of the study presented emphasize another question raised during the 2010 SAP related to pseudo-steady state versus spike exposure. If the effects of atrazine on the adrenal play a role in effects of atrazine on the LH surge, it is not clear if the overall effects on the LH surge are due to attaining systemic steady state levels of atrazine or to the continued exposure to elevated levels of progesterone occurring with repeated dosing.

The overall conclusion that a single isolated exposure to atrazine is not likely to elicit adverse reproductive outcomes is made with the underlying assumption that atrazine exposure is affecting only the LH surge. Further, long term exposure where dynamics of the LH surge might be altered over an extended period of time (many months) or for an extended period of time during different stages of the reproductive lifespan (puberty, menopause) may potentially elicit adverse outcomes.

Charge Question 7 - An increased incidence of prostatitis has been observed in the offspring of rats exposed to atrazine from PND 1 to 4. This effect was not linked to an LH alteration, but rather the atrazine related suppression of suckling-induced prolactin release in the lactating dams. Prolactin plays a crucial role in the neonatal brain for normal TIDA neuron development. In the adult offspring, the impaired TIDA regulation is reflected by elevated prolactin levels (hyperprolactinemia). It is this elevated level of circulating prolactin in the adult male rats that has been linked to an increased incidence of prostatitis. It is unknown when the TIDA neurons develop in the human fetus or whether this development is dependent on the maternal prolactin concentrations.

a) Given the biological processes involved in the atrazine-mediated prostatitis in rats, please comment on the human relevance of these findings in rats for the overall hazard characterization for atrazine.

Panel Response:

Prostatitis is the most common urological diagnosis in men younger than 50 years of age and the third most common urologic diagnosis in men older than 50. Approximately 50% of all men experience prostatitis-like symptoms at some point during their lifetime. The National Institutes of Health has re-categorized the types of prostatitis into four types: acute, chronic, chronic pelvic pain syndrome, and asymptomatic histologically-evident inflammation.

Lewis, Copenhagen and Wistar rats develop a spontaneous nonbacterial chronic prostatitis (CP) with advancing age, making them good animal models for laboratory investigation of this disease (Naslund, 1988; Robinette, 1988). Furthermore, a number of studies have demonstrated that spontaneous nonbacterial prostatitis in rats was histologically very similar to CP in humans (Muntzing et al., 1979).

Treatment of male rats with estradiol (E2) increases the incidence and severity of inflammation (Seethalakshmi et al., 1996; Naslund et al., 1988). Estrogen-induced prostatitis is partly related to the inhibition of dopamine secretion at the hypothalamus, and dopamine deficiency enhances the production and secretion of prolactin that eventually causes inflammation of the prostate (Tangbanluekal and Robinette, 1993). These authors reported that E2-induced prostatitis in rats was correlated with increased serum prolactin, and elevated pituitary weight, and that the administration of bromocryptine, a dopamine D2 agonist, was effective in suppressing pituitary weight and hyperprolactinemia and mitigated the inflammation observed in the lateral prostate region of the gland.

A few studies have indicated that exposure to atrazine during late gestation (GD 15-19) or the early postnatal period (PND 1-4) in male rats leads to inflammation of the prostate (ventral and lateral regions) at later stages of growth (PND 90-120). Elevated prolactin levels are thought to play a significant role (Tangbanluekal & Robinette, 1993; Stoker et al., 1999; Rayner et al., 2007; Stanko et al., 2010). The mechanisms underlying the cause of the prostatitis are not yet defined, but may be related to changes in the hormonal milieu during critical periods of development, which, in turn, may have adverse effects during aging. It is important to note that

all these studies have used short-term exposure with a range of doses and different rodent species.

With regard to human relevance:

• It is unlikely that actual human exposure would follow a similar brief pattern employed in the animal studies. Human exposure to atrazine and its degradates is more likely to occur over a lifetime and at lower levels than those used in the animal studies. *In vitro* and *in vivo* studies, in human and animal tissues, evaluating cellular and molecular expression changes in response to atrazine at comparable human exposure levels should be conducted.

• The prostate of the rodent (both mouse and rat) consists of distinct lobes (unlike the human) and has been the animal of choice for reproductive biology research for many years. One of the values of this model is the ability to examine the effects of treatments, such as endocrine disruptors, and observe which region or regions have a specific sensitivity to the compound under investigation. Based upon such studies, it has become clear that the region of the rodent prostate exhibiting the most sensitive response to the effects of estrogenic endocrine disruptors was the dorsolateral lobe (Timms et al., 2005). While not the primary objective of the atrazine studies cited in the EPA Issue Paper, there is some evidence that atrazine effects are mediated through alterations of steroidogenesis, including estrogens. It is important to emphasize that the dorsolateral prostate in the rodent animal models is homologous to the major prostate region in the human male (the peripheral zone – Timms, 2008; Final EPA Issue Paper, p 30). While most of the studies have looked at prostatitis in the lateral lobe, the dorsal lobe or combined dorsolateral lobe typically have not been examined. Of interest was the observation from an earlier anatomical study (Timms et.al., 1994) that this region was homologous to the peripheral zone of the human prostate, the region most susceptible to the development of cancer.

• In the atrazine rodent studies, tissue weights are measured to determine growth and/or adverse effects. In a recent National Toxicology Program scientific review of an endocrine disruptor (NTP, 2008), a specific concern was raised regarding the reliability and usefulness of this variable as a measurement of physiological effect. As stated in that report: “Perhaps the most important confounding factor in all of the prostate studies is that prostatic wet weight is an extremely poor measure of prostatic growth, which substantially diminishes the strength of data advanced both for and against an effect … on prostatic growth.” In light of this, the inflammatory response should be characterized in this model using contemporary approaches such as BrdU labeling and immunocytochemistry to determine proliferation rates, and characterization and histological quantification of the hyperplastic response. The use of tissue microarray following laser capture micro-dissection might better define the cellular mechanisms responsible for the region-specific inflammatory responses. Based on several animal models, inflammatory mediator induction would likely occur prior to, and at lower doses than, histologically evident cellular inflammation. Cytokine induction can induce a number of effects on the tissue microenvironment including hyperplasia, desmoplasia, and dysplasia. Furthermore, cytokines induce developmental growth regulators including IGF-1, TGFb, and FGF. Levels of these factors should be evaluated. Finally, inflammation is associated with DNA damage and loss of imprinting of some genes. This should be evaluated as part of the characterization.

• One interesting aspect of this gland in the rodent, not discussed in any of the relevant publications, is the fact that the prostate, especially the lateral lobe, has a high level of endogenous zinc. It may also be relevant that tissue zinc levels are reported to be lower in prostate pathology (benign prostatic hyperplasia (BPH and cancer), which raises an interesting question as to whether dorsolateral levels of zinc are reduced in early prostate development by atrazine exposure and whether this plays a role in the later development of prostatitis. The relevance of this to the incidence of prostatitis may be of importance with regard to mediators of inflammation. Zinc concentrations and their correlations to inflammatory mediator expression and cellular inflammation should be evaluated.

• To date, there has been no conclusive causal link made between atrazine exposure and prostate cancer. However, several important questions remain unresolved from the present literature, such as the effects of repeated life-long exposure to low doses on both prostate cancer incidence and grade, and any effects on the progression to advanced disease. The causes of prostate cancer are unclear, but several studies have indicated that chronic prostatic inflammation may precede benign prostatic hyperplasia and/or cancer in humans, depending on the zone. In fact, inflammation is the most tightly correlated histological anomaly to prostate cancer. If the link between prostatic inflammation and atrazine exposure is confirmed, this ascends to added importance and any related consequences should be considered.

Treatment of Wistar dams with daily doses of atrazine on postnatal days (PND) 1-4 resulted in suppression of suckling-induced prolactin (PRL) release in offspring. At 120 days of age, male offspring showed increased incidence and severity of prostate inflammation in the ventral and lateral lobes (Stoker et al., 1999). Except for a significant increase in ventral prostate tissue weight in the lowest dose (6.25mg/kg – calculated as non-significant as a percentage of body weight), no other weight changes were observed when compared with controls. The lateral lobes were examined by myeloperoxidase assay and histology. The type of inflammation was characterized as a focal neutrophil infiltrate in the lumen and focal mononuclear cells in the stroma. Although described as a chronic inflammatory response, the authors did not make a comparison of the inflammation with the classification of human prostatitis types (bacterial/non-bacterial) (Barry and McNaughton-Collins, 2007).

• A major point in considering the relevance of this study to human health hazard is with regard to the mechanism(s) of action for the exposure - offspring suckled from atrazine-treated dams, and at PND120 the lateral prostate (right lobe) was evaluated for evidence of prostatitis. A recent report card from the Centers for Disease Control (CDC) indicates that 3 out of 4 mothers in the US initiate a breast feeding regimen, but by 3-6 months, this rate is not maintained (CDC, 2011). The study by Stoker et al. (1999) suggests that early lactational exposure to prolactin is important for the normal development of tuberoinfundibular neurons. However, according to the CDC data, approximately 25% of US babies are not exposed to breast milk prolactin, which may pose the question as to whether these individuals are at the same risk for later development of prostatitis?

• What are the estimated or measured levels of atrazine and its metabolites in human breast milk?

• The baby boomers (i.e., those born between1946 and 1964) and subsequent generations are most likely to be affected by any health hazards from atrazine exposure, given that the chemical has been in use only since the 1950s. It is imperative to determine the actual exposure levels of the human male and female population in the US, especially where there may be an increased risk of exposure. Determination of the actual exposure levels will better define the correlative experimental exposure levels and frequency.

Several unresolved questions limit the conclusions that can be drawn regarding the human relevance of atrazine's effects observed in the rodent prostate. Results in rodent models are of limited applicability due to inconsistencies in atrazine exposure levels and methodology. Secondly, in rodents the inflammation caused by atrazine has not been sufficiently characterized regarding molecular and cellular events that may indicate critical changes to the tissue microenvironment, leaving open the possibility that lower doses of atrazine could produce subtle but biologically significant effects. The cellular signaling mechanisms involved have not been elucidated and molecular events such as DNA damage and imprinting changes may be possible at low levels; such changes may accumulate during the aging process of men. Finally, it is unclear what effects atrazine may have on the truly relevant measures of prostate cancer effects, including cancer grade, progression, and aggressiveness.

It is important to emphasize that the data regarding atrazine and effects on the prostate remain inconclusive, both in human and in rat models. Studies that have investigated prostate cancer risks in plant workers exposed to atrazine have concluded that detection bias accounts for any increased risk in exposed individuals. This may be true; however, with the currently available limited studies (i.e., only 5 cases investigated and with no investigation into disease grade, progression to androgen-refractory forms, or age of diagnosis), it is not possible to tell if atrazine increases prostate cancer risk. Prostate cancer affects nearly half of American men, but the majority of afflicted individuals exhibit a clinically insignificant form represented by Gleason scores of 3+3 or less. Given the prevalence of this form of the disease, and the clinical insignificance of its pathology, there will not be statistical power to determine if atrazine or any other agent increases risk, if overall lifetime incidence is the measure. However, given the importance of the aggressive forms of this disease, characterized by Gleason scores of 4+3 or higher and androgen-refractory growth, these measures should be analyzed in exposed individuals to determine if atrazine truly has an effect on prostate cancer risk. The grade of the disease, Gleason score, and progression to metastasized forms and androgen-independence should be the endpoints analyzed by future epidemiology studies.

The prostatitis studies performed in rats are also inconclusive due to the irrelevance of the extremely high doses used in the study, relative to actual human exposure. Future mechanistic studies should be aimed at evaluating the effects in an appropriate model with doses relevant to human exposure, whatever those are determined to be. This not only includes the concentration, but also the method, as humans are exposed to repeated and continual concentrations of atrazine, not one single high dose. The studies are also inconclusive in terms of relevance to human health because the inflammation has not been carefully characterized. As discussed earlier, more subtle effects such as cytokine expression patterns and genetic imprinting should be investigated. It is possible that important changes to the prostate may occur at lower doses that do not result in histologically evident cellular inflammation. This is important because subtle molecular changes

could have serious effects on tissue microenvironments. If carefully-conducted dose-relevant studies determine that atrazine induces prostatic inflammation at relevant doses, the effects on human prostate biology could be profound. Inflammation is the most associated histologically evident factor associated with prostate cancer (Nelson et al., 2003; De Marzo et al., 2007). It has been postulated by many prostate biologists that inflammation may induce or promote prostate cancer, similar to what has been observed in colon and ovarian cancers. A wealth of literature support that inflammation induces putative pre-neoplastic lesions, including proliferative inflammatory atrophy (PIA) and prostatic intraepithelial neoplasia (PIN) (De Marzo et al., 1999). Further, inflammatory mediators induce prostate cell growth in culture, and induce growth factors in tissue culture (Begley et al., 2008). Inflammation of the prostate is also associated with increased DNA damage and lower-urinary tract symptoms (LUTS) associated with benign prostatic hyperplasia (Steiner et al., 2003; Nickel et al., 2007; Nickel et al., 2008). At a recent Society of Basic Urological Research meeting (Spring 2011), a panel concluded that the most commonly associated finding in men with significant clinical LUTS is inflammation—much more so than prostate size (Nickel et al., 2007; Nickel et al., 2008). LUTS has not been investigated in response to atrazine exposure, and although not a deadly condition, BPH-associated LUTS has enormous impact on quality of life in American men. It is very likely that, if relevant atrazine human exposure does induce inflammation in humans, that this would result in significant clinical implications.

The causes of inflammation in the human prostate have not been elucidated. It is not likely to be caused by bacterial infection in most cases, since culturable and non-culturable (16S rRNA) studies have not produced conclusive results despite large study groups. Possibilities that remain include viruses, systemic changes resulting from obesity, type 2-diabetes and metabolic syndrome, dietary factors and exposure to environmental factors including endocrine disruptors (reviewed in De Marzo et al., 2007). Epidemiology has further revealed that the lifetime prevalence of prostate cancer in American men is nearly 50%, but less than 10% in Asian countries. In China the prevalence is 1.5%. Chinese-Americans that live in the United States for two generations exhibit incidences nearly equal to those of their Caucasian neighbors. Given this, the role of inflammation in prostate cancer and the clear clinical relevance of prostatic inflammation to public health, the Panel recommended that any possible role atrazine may have in the etiology of prostatic inflammation be investigated with carefully-controlled dose and model-relevant studies.

Charge Question 8 - When evaluating the data on mammary gland development, the September 2010 SAP report concluded that the "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." Regarding the methodological differences between the Rayner et al., and the Coder studies, the Panel commented that "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." In response to this comment, the Agency has conducted a set of experiments investigating the potential impact of in utero atrazine exposure on mammary gland development (MGD) in Sprague Dawley rats using both the subjective scoring methodology described by Rayner et al. and a computer-based quantitative methodology (morphometric analysis). Using either quantitative or subjective measures to evaluate mammary gland development, no differences in MGD were found

between control or atrazine-exposed rats indicating that gestational atrazine exposure had no demonstrable effect on normal mammary gland development.

a) Please comment on the Agency's findings in addressing the issues raised by the SAP during the September 2010 meeting. Please comment on whether this study (along with the negative studies by Coder) adds to the weight of evidence that it is unlikely that atrazine impacts mammary gland development.

Panel Response:

The new data presented by EPA ORD NHEERL addressed the concern that the studies from Fenton's group and those from Hovey (Syngenta) used quite different approaches for capturing mammary gland morphology. In the Fenton group's study, a ranking system was used, and, in the Hovey study, a set of measured morphometric variables was applied. While the ranking system has been referred to in the EPA Issue Paper as "subjective" and "qualitative", it is in fact, neither. When done in a blinded fashion with trained application of morphological criteria, the method is clearly "objective", and no less so than a morphometric approach. It is also quantitative because the established morphological criteria are converted to quantities, i.e., ranks that can then be compared with standard statistical methods. Morphometric measurements as used in the Hovey studies may be objective, but only if the measured variables are not chosen subjectively. The 2009 workshop referenced in the EPA Issue Paper did a good job of summarizing best practices for using morphological variables to characterize rodent mammary glands.

Applying both approaches (ranking and morphometry) in a careful manner with SD rats, the NHEERL study presented in the EPA Issue Paper demonstrates that: 1) both approaches produce similar conclusions and 2) any effects of prenatal atrazine exposure on mammary gland development early in life are very subtle and not measurable by either of these methods. While Long-Evans rats might be suggested as a different choice of animal model, good arguments have been made that the SD strain is appropriate and adequate.

The use of mammary gland "development" as an index of adverse environmental chemical effects has been advocated based on a number of features of mammary gland growth, morphology, pharmacology, and physiology. These features include the exquisite hormone-responsiveness of the mammary glands and their distinct developmental sequence of events, most of which occur after birth. These are compelling notions, but thus far implementing toxicological studies using mammary gland endpoints has been difficult. Given the centrality of lactation in the life history of mammals, continued concern about the effects of environmental chemicals on mammary gland biology is extremely important. Therefore, inadequacies in the literature relating to atrazine effects on mammary gland development should not deter future studies; in fact, they should encourage them to be done.

The direct answer to the charge question is that the new evidence does not provide any support for an effect of atrazine on mammary gland morphology. In the new studies, SD rats were treated *in utero* with a wide range of atrazine doses, and tissues were taken for analysis on postnatal day 45, when pubertal mammary gland development is presumably complete.

Mammary gland morphology was measured by an arbitrary rank scoring system, and by a morphometric quantification system using image analysis. Because certain morphological characteristics occur in a predictable manner, measurements of the sort presented here are taken to signify "development." However, it is important to remember that development refers to the processes that underlie morphological change, not the morphology, per se.

While the Charge Question is focused on resolving differences in experimental design, data gathering, and interpretation between studies from the Fenton lab (Raynor, et al., 2004; 2005; Enoch, et al., 2007), and those from NHEERL (Davis, et al., 2011), and Hovey (2011), these ambiguous findings need to be considered within a larger context.

Earlier concerns about breast carcinogenesis in atrazine-treated rodents were resolved satisfactorily by discovering that mammary tumors came about by a process of disordered postnatal development driven by accelerated reproductive senescence and inappropriate secretion of gonadotropins, steroids and prolactin (Cooper et al., 2007). These well-accepted conclusions lead to the simple deduction that atrazine does have effects on mammary gland development, even if those effects do not appear unambiguously in the results from the early life studies from Fenton's lab and the other studies cited here.

The more relevant question seems to be whether development of the rodent mammary glands early in life provides an adequately robust model in which to observe subtle adverse effects of a potential environmental toxicant, such as atrazine. For a variety of reasons, it seems unlikely that mammary gland morphology, standing as a surrogate for underlying developmental processes, is adequately sensitive to fulfill this role. One limitation of rodent mammary gland morphology is that it is subject to wide variations among rodent strains depending on differences in hormone secretion patterns, the presence of endogenous retrovirus (particularly Mouse Mammary Tumor Virus), and nutrition. In addition, there are internal differences between morphological characteristics of the glands within an individual and within regions of any particular gland. Another limitation is one's ability to define differences in morphology ("development") as being adverse. Given that the function of the glands is to produce adequate milk for the offspring, for any change in morphology to be defined as adverse, it would need to be connected in some objective way to a deficiency in milk supply. Given that the glands are controlled by a host of intrinsic and extrinsic homeostatic mechanisms that regulate milk production, it is not surprising that subtle effects of environmental chemicals may not, by themselves, perturb mammary gland function sufficiently to be definitively "adverse."

Concerns remain, however, as to whether an environmental toxicant such as atrazine, which affects reproductive hormones or other mammary-related physiological variables, might interact in important ways with other environmental factors that predispose individuals to poor mammary gland function. In particular, obesity is a known risk factor for poor mammary gland function in humans, as well as in rodent models. It is certainly conceivable, maybe likely, that subtle effects of an environmental chemical will have important consequences in overweight individuals.

In conclusion, it is true that the current data "adds to the weight-of-evidence that it is unlikely that atrazine impacts mammary gland development." However, the evidence of effects

in some studies, combined with the known effects of atrazine on reproductive hormones, provides an important basis for continued concern and efforts to design better studies that would determine whether these hormonal effects could contribute to poor lactation, a clearly adverse outcome, in susceptible individuals.

Charge Question 9 - In 2003, an FQPA Safety Factor was retained in part to address the lack of data evaluating the potential for differences in sensitivity across life stages. Since then, multiple studies evaluating atrazine's potential impact following exposure during the prenatal, perinatal, prepubertal, and adult stages of life have been conducted to address these data gaps. None of the new studies has identified an enhanced sensitivity in the young and the LH surge attenuation observed in females of reproductive age continues to be the most sensitive endpoint.

a) Please comment on the weight of the evidence analysis conducted by the Agency and the extent to which the uncertainties related to the potential for differential sensitivity of the young are addressed with the additional data.

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 manner consistent with that for other pesticides. This includes, but is not limited to, the extent that the database on hazard and mode of action informs us about potential increased susceptibility of infants and children. The selection of the FQPA safety factor to be applied in a given situation is predicted on the basis of judgment as to whether or not sufficient, credible scientific data are available to determine if pre- and post-natal toxicity could occur at exposures less than those associated with induction of critical effects (in this case, attenuation of the LH surge) or if available data are inadequate as a basis for characterizing exposure. Application of uncertainty or safety factors should also take into account the relative degree of protection offered by the selected point of departure, in this case, an early precursor event (33% LH surge attenuation) in a sensitive strain rather than an adverse effect.

The Panel considered this Charge Question in the context of the following components:

1) Does the existing body of data exploring the potential for adverse consequences following exposure (direct or indirect) encompass all of the life stages of interest (i.e., prenatal, perinatal, pre/peri-pubertal, and adult)?

2) Do the currently available studies allow for an adequate assessment of the potential for differential sensitivities, in light of the fact that some of these studies do not include measurement of the same endpoint/phenomenon that currently serves as the basis for the quantitative characterization of hazard (suppression of the LH surge)?

3) Does Chapter 5 of the EPA Issue Paper (Scientific Considerations in Potential Sensitivity of Infants and Children) adequately present WOE and uncertainty analyses on the question of potential age-related differences in sensitivity?

Weight-of-evidence analysis of the existing data which bear on the determination of the potential for life stage-related differences in sensitivity is not evident. The summaries of the studies that became available since the last SAP meeting are too brief to enable meaningful evaluation. A truly robust WOE discussion would include a recapitulation of critical information on the previously-submitted/reviewed studies along with the newer ones, as a basis for consideration of the adequacy of the database. It would also include discussion of the extent to which the uncertainties in the database related to determining the potential for differential sensitivity of the young are addressed.

It would have been helpful/enlightening to have had presented in the EPA Issue Paper, a figure or table which summarized all of the relevant studies bearing on this issue. EPA included several tables in its presentation on the Scientific Considerations in Potential Sensitivities of Infants & Children. The figure on Slide #9 of the Syngenta presentation on mode of action also represents a possible approach to data summarization. These could serve as starting points for a composite table/figure of the studies.

Assessment for weight-of-evidence for potential for differential sensitivity should include consideration of consistency, dose-response concordance, specificity and biological plausibility. Biological plausibility should take into consideration weight-of-evidence for hypothesized mode(s) of action (MOA).

Based upon its own review and assessment of the existing database as presented in the current EPA Issue Paper and that available for discussion at the earlier SAP meetings, the Panel has concluded that there is sufficient information available to reach the conclusion that the issue of differential sensitivity has been adequately studied. This relatively extensive database, spanning all life stages from conception to adulthood indicates no unique susceptibility to atrazine in the developing organism. The Panel agreed with the Agency's conclusion that exposure during the earlier life stages does not lead to greater sensitivity, if one accepts the premise that the data on suppression of the LH surge is appropriate for use in making the comparisons.

Additionally, the proposed point of departure, based upon attenuation of the LH surge, appears to be protective against adverse reproductive/developmental outcomes such as delays in onset of puberty, disruption of ovarian cyclicity and inhibition of suckling-induced prolactin release. Given the interplay between the point of departure and applied uncertainty or safety factors, in the interest of transparency, it would also be helpful to consider an array of points of departure for various endpoints and their biological significance, with a view to bounding potentially the degree of conservatism associated with the ultimate choice. This would include, but not be limited to, the BMD/L for the impact on the LH surge, but including also those for more traditional apical endpoints, generally considered to be adverse. This seems rather critical as a basis to interpret the derived BMD in the context not only of its biological significance but

also its degree of conservatism in the risk assessment construct which has traditionally been used to address more severe endpoints.

Even though aspects relevant to exposure are still evolving, the Panel wondered, additionally, if any thought has been given to estimating the internal dose metric for younger age groups of the human population (intraspecies variation) for consideration in the context of FQPA, recognizing that the chemical specific adjustment factors for interspecies (animal to human) kinetic differences are likely to be less than default.

As summarized in EPA's policy guidance entitled, Determination of the Appropriate FQPA Safety Factor(s) in Tolerance Assessment, Section 408 (b) (2) (c) of FQPA "instructs EPA, in making its "reasonable certainty of no harm" finding, that in "the case of threshold effects, ...an additional tenfold margin of safety for the pesticide chemical residue and other sources of exposure shall be applied for infants and children to take into account potential pre- and postnatal toxicity and completeness of data with respect to exposure and toxicity to infants and children." Section 408 further states that "the Administrator may use a different margin of safety for the pesticide chemical residue only if, on the basis of reliable data, such margin will be safe for infants and children."

It should be noted that the law does not impose any directional constraints on the choices for "a different margin of safety." The "different margin of safety" could be greater than 10X or less than 10X. While this flexibility exists in the law, there is little precedent, explicitly or implicitly, for application of an FQPA Safety Factor greater than 10X, substantial precedent for the reduction of the FQPA Safety Factor to 3X or 1X, but no precedent for application of an FQPA Safety Factor of less than 1X. (Cases where the FQPA Safety Factor has been "removed" equates to an FQPA Safety Factor of 1X).

As articulated in the 2003/2006 RED, EPA retained "the FQPA Safety Factor of 10X for atrazine and its chlorinated metabolites to protect the safety of infants and children in assessing risk from dietary (food and drinking water) exposures. The FQPA safety factor was reduced to 3X for residential exposure. EPA stated that "this is considered adequate to protect the safety of infants and children in assessing residential exposure and risks because the exposure concerns for drinking water included in the 10X FQPA safety factor for dietary exposure do not apply to residential exposure scenarios, although the concerns for the effect of the neuroendocrine mode of action on the development of the young remain." The July 2011 version of the "Re-evaluation of Human Health Effects of Atrazine…" summarizes the results of a series of studies (some pre-dating, some post-dating the Sept 2010 SAP meeting) and concludes: "Although additional experimental toxicology studies are still on-going to better characterize the potential adverse health outcomes resulting from atrazine exposure (including the duration of exposure that may lead to an adverse health outcome), the available data do not indicate that pre- and/or post-natal exposure leads to increased sensitivity in the young relative to the attenuation of the LH surge that serves as the basis for the atrazine risk assessment."

The current chronic reference dose of 0.018 mg/kg/day (and PAD of 0.0018 mg/kg/day) is based upon the results of the 6 month study (Morseth et al., 1996) which established a lowest observed adverse effect level (LOAEL) of 3.65 mg/kg/day for attenuation of the LH surge and a

NOAEL of 1.8 mg/kg/day. The NOAEL from this study was the basis for the point of departure for the chronic reference dose (cRfD). In the July 2011 EPA Issue Paper, the Agency proposes to replace the 1.8 mg/kg/day NOEL from the Morseth study with the 2.56 mg/kg/day BMDL derived from the LH surge data in the Cooper et al., 2010 study, to serve as the PoD for deriving a cRfD and PAD. EPA goes on to note that "Three additional studies evaluating the effect of atrazine exposure across life stages have become available within the last few months. These studies reinforce the conclusions reached during the September SAP meeting since all of the effects observed in the young in these [sic] set of studies occurred at doses ≈ 25 times higher than the dose EPA is proposing to use (BMDL of 2.56 mg/kg/day) as the PoD for human health risk assessment (derived from LH data collected after four days of exposure to adult females)."

Given the apparent lack of early age-related sensitivity to the neuroendocrine effects that are driving the hazard assessment, several options with regard to the appropriate magnitude of the FQPA safety factor could be considered.

a) Option 1: The 10X FQPA safety factor currently applied in the dietary risk assessment could be reduced to 3X, removing the 3X/reducing to 1X that portion of the FQPA safety factor which addresses the concerns regarding the hazard potential. The 3X FQPA safety factor currently applied in the residential risk assessment also would be removed/reduced to 1X. (Of course, when the drinking water exposure issues are resolved, a re-visit of the remaining 3X component of the FQPA safety factor applied in the dietary risk assessment should occur).

b) Option 2: The conclusion that one could draw from the Agency's statement above is that not only is there no differential increase in sensitivity in the young as the consequence of pre- and/or (early) postnatal exposure, there is, in fact, a decreased sensitivity when compared with the adult female. On the basis of this finding, one might argue that the 3X FQPA Safety Factor currently applied to account for the uncertainties/concerns around the neuroendocrine effects could be reduced to less than 1X. For example, on Page 29 of the 2011 EPA Issue paper, the Agency states that "All the effects on sexual maturation and altered androgen status reported after ≈ 30 days of exposure occur at dose levels ≥ 5X higher than those leading to the LH surge disruption which serves as the basis of the Agency's current risk assessment" and "The dose level eliciting the increase in the incidence in prostatitis in the offspring is > 10-fold higher than the dose leading to the LH surge attenuation used as the basis for the Agency's risk assessment." This argues that the FQPA Safety Factor component addressing the hazard potential could be reduced not just to 1X, but further by at least five-fold (i.e., to 0.2X or less).

c) Option 3: Consistent with the idea posited above that it would be helpful to consider an array of points of departure for various endpoints and their biological significance, the Agency could replace the NOAELs/LOAELs for suppression of the LH surge with NOAELs/LOAELS for actual adverse outcomes/endpoints in adults exposed only as adults, then compare the latter with NOAELs/LOAELs observed in studies which evaluate apical endpoints in animals exposed during earlier life stages. It would be of value to know if the same lack of apparent differential sensitivity still holds. In this instance, one would be comparing apples to apples (i.e., an apical adverse outcome to an apical adverse effect) as opposed to apples to oranges (i.e., a non-adverse precursor effect to an adverse effect).

Charge Question 10 - Section 3.0 and Appendices B-3 through B-5 of the draft Issue Paper provide the Agency's reviews and synthesis of the atrazine cancer epidemiology studies. These studies examine a variety of cancer endpoints, notably reproductive and endocrine system tumors including prostate, breast, ovarian and thyroid tumors, cancers of the lymphohematopoietic system including non-Hodgkin's lymphoma (NHL), and other cancer types including glioma, pediatric cancers and colon cancer. Section 3.3 integrates the findings of the epidemiology and experimental toxicology studies.

a) Please comment on the sufficiency of the Agency's cancer epidemiology reviews with respect to identifying the major strengths and limitations of each study, and overall synthesis of results by cancer types.

Panel Response:

Overall, the Panel found EPA's review of the epidemiologic evidence to be comprehensive and systematic. The framework used was appropriate, and broad selection of literature as opposed to using a rating system allowed for a thorough review of the epidemiologic evidence. Each study was reviewed using the framework with strengths and limitations appropriately noted for the most part. The overall synthesis of results seems appropriate for prostate, breast, ovarian, and other cancers; however, no statement was made as to the likelihood of an association for thyroid cancer, although the strengths and limitations of the study in which it was evaluated are noted. Moreover, the Panel believed that the classification, "not likely to be carcinogenic to humans" overstates the epidemiologic evidence, as well as the "weight of the evidence," against a causal association between atrazine and cancers in humans. Specific additional strengths and weaknesses to be considered are noted below.

Literature Search

EPA's literature search methods were sufficiently thorough and transparent as described in Appendix B, pages 17-18. However, the description of the literature search methods in section 3.1.2 of the EPA Issue Paper is too brief. It would be beneficial if more of the description provided in Appendix B were to be included in the EPA Issue Paper.

The Agency describes the inclusion criteria for their review as "purposely broad and reflected only that atrazine or triazine is measured in association with a cancer outcome." However, the selection process for studies to include for review only involved review of the title and abstract, which actually is a narrow, not broad, approach to inclusion because findings with regard to the outcomes could appear in the text of papers that were not included as a consequence of the findings not being mentioned in the key words, title or abstract. Thus, both the lack of provision of search terms and only reviewing title and abstract for eligibility may have resulted in some studies with relevant findings being missed. Although the Panel believed that EPA captured all of the relevant epidemiologic studies on atrazine or triazines and cancers, for future literature searches the Panel urged EPA to check not only the title and abstract, but also the article text and result tables to determine if associations between a cancer site and atrazine or triazines are evaluated.

Literature Review Methodology

EPA's method of evaluating the epidemiologic studies was, in general, sufficiently comprehensive. Important aspects of study design were considered, including accurately measuring the cancers and exposures, issues of bias, sample size and statistical power. Major strengths and weaknesses were identified, in particular whether exposures were assessed quantitatively, the ranges of exposure, and whether critical exposure time windows were evaluated. Established etiological criteria were used in the evaluation, including temporality, magnitude of the measure of association (e.g., relative risk, odds ratio), exposure-response trend, consistency of findings across studies, and biological plausibility.

However, the Panel had some issues with EPA's methods of assessment. First, the Panel believed that the focus of the assessment should be on individual-level studies. Ecologic studies should be evaluated only if there are compelling reasons to do so; e.g., if there are no individual-level studies for the cancer site. Moreover, because of the limitations of ecological studies for etiological inference, negative findings in ecological studies should not be used to counter positive findings in individual-level studies.

Second, the focus of the assessment should not be on whether a finding is "statistically significant," given the low statistical power of most of these studies. Emphasizing statistical significance when statistical power is low will likely result in committing type 2 errors. It should be kept in mind that a finding may be etiologically significant (and may also have public health significance) even though it is not statistically significant. Moreover, the term, "borderline significance" is used in several places in Appendix B and should be avoided because a finding is either statistically significant or it is not. Findings that are slightly below or above the level of statistical significance, (e.g., a p-value of 0.05), do not have appreciably different likelihoods of being chance findings.

Third, concerning biases, it is important to provide some evidence that a specific bias is likely to be present as well as the likely magnitude and direction of the bias. For example, it is not helpful to simply state that confounding by other pesticides is a limitation in a study without providing any evidence that these other pesticides are actually confounders. It is not enough for a pesticide to be highly correlated with atrazine; to be a significant confounder, the pesticide must also be a strong risk factor for the disease of interest. As one review article has stated: "…even for studies of occupational exposures and lung cancer, tobacco-adjusted relative risks rarely differ appreciably from the unadjusted estimates" (Blair et al., 2007). Even though smoking is highly prevalent in occupational cohorts, and it is an extremely strong risk factor for lung cancer, tobacco-adjusted relative risks for the association between an occupational exposure and lung cancer are not likely to differ from unadjusted estimates by more than 20%. However, the same review concluded that a small amount of non-differential exposure misclassification bias can have a profound impact on measures of association (Blair et al., 2007). For example, even if there is a relatively good correlation (e.g., r=0.4) between the AHS exposure intensity score and urinary levels of atrazine, the resulting non-differential exposure misclassification can produce an observed relative risk (RR) that is half of the true RR (Blair et al., 2011). In this scenario, if the true RR was 2.0, the observed RR would be ≤ 1.4 (Blair et al., 2011). The likely

impacts of non-differential exposure misclassification should be taken seriously – i.e., considerable bias toward the null for dichotomous exposure variables, and distortion of exposure-response relationships so that monotonic trends are not observed. (Healthy worker survivor effect biases also have a similar impact on the exposure-response relationship – attenuation of effects at the higher range of exposures so that a monotonic trend is not observed.)

Fourth, the Panel took issue with the following statement from page 71 of the EPA Issue Paper: "the weight of the evidence supports that atrazine is not likely to be carcinogenic in the human population." First of all, there is considerable uncertainty and gaps in the toxicological evidence concerning whether atrazine is a human carcinogen. Second, EPA has not done a comprehensive "weight of the evidence" assessment; instead, the toxicological evidence appears to be used to nullify any positive evidence from the epidemiologic studies. Third, the evidence across cancer sites is mixed, not uniform, with some cancer sites having no evidence for an association, whereas other cancer sites having at least suggestive evidence for a causal association. Even though the Panel agreed with EPA that the epidemiologic evidence does not strongly suggest a link between atrazine and specific cancers, the Panel did not agree that a lack of strong evidence justifies a conclusion that atrazine is not likely to be a human carcinogen. EPA does acknowledge that "some epidemiology studies are weakly suggestive of an association between atrazine exposure and cancer incidence in the human population" (p. 71). The statement that atrazine is not likely to be carcinogenic in humans ignores this evidence.

Finally, although the EPA Issue Paper appropriately summarizes the epidemiologic evidence for each cancer site, the appendix is disorganized and hard to follow because it does not evaluate the evidence for each cancer site separately. The table of studies in the appendix is also disorganized in a similar fashion. The Panel recommended that EPA reorder the appendix by cancer site, and create a table separately for individual-level studies that is ordered by cancer site. The table should include all the individual-level studies that have been conducted, including those prior to 2004.

Agricultural Health Study (AHS)

The Panel found that EPA summarized in a generally comprehensive fashion the strengths and limitations of the Agricultural Health Study (AHS). One significant strength that is not mentioned is the longitudinal follow-up with exposure and risk factor information updated every 5 years. The EPA Issue Paper does not indicate if the analyses of cancer included repeated measures, particularly for exposure assessment, and also for time-varying covariates, which, if done, would be a significant strength to add because of improved exposure assessment and control of covariates by use of repeated measures. The exposure intensity index used in the AHS largely addressed concern about not using repeated measures for exposures, but the concern about adequate adjustment for covariates by using repeated measures remains and is noted subsequently as a possible additional limitation. Although the updated information has not been included in the AHS analyses so far (most likely because of latency considerations since the phase 1 information was obtained in the mid-1990s and data on cancers have been available only up to 2007), eventually the AHS will be able to utilize this information to create time-varying variables for atrazine exposures and risk factor covariates. The EPA Issue paper mentions that a strength of the AHS is that it collects information on family history of prostate cancer; but

information is also collected for family history of other cancers as well, including breast cancer (Engel et al., 2005).

With regard specifically to the AHS, five major limitations that were not emphasized should be highlighted. First, the criteria for diagnosis and coding of cancer outcomes were not necessarily standardized across the physicians and hospitals in their reporting to the cancer registries, which could potentially result in misclassification of the outcomes. Second, the AHS study population is predominantly "white," which is not the highest risk group for incidence or mortality of some cancer outcomes; e.g., prostate cancer incidence and breast cancer mortality. Thus, this study cannot determine if such high risk, "non-white" groups might have particular susceptibility to (or protection from) atrazine-related exposures, or if the findings are generalizable to these high risk groups. Third, the number of women who are occupationally exposed to pesticides in the AHS cohorts is relatively small, limiting the ability (e.g., because of very low statistical power) of these studies to evaluate cancers and etiologically-heterogeneous cancer subtypes among female pesticide applicators, or to effectively adjust for possible confounding. Fourth, while it is understandable that adjustment would be made for use of other pesticides that are highly correlated with atrazine use, if these pesticides are not independent risk factors for the disease of interest, then their inclusion in regression models could produce unstable results and would also tend to attenuate the effect measure for atrazine. Fifth, the EPA Issue Paper notes in several places that selection bias was unlikely to play an important role in this study because loss to follow up was only 2%. However, loss to follow up is not the only source of potential selection bias. Another major source of selection bias is the initial participation rate. The AHS managed to enroll 82% of its target population, which is good; but the 18% non-participation rate also contributes to the potential for selection bias. In addition, healthy worker biases could be considered sources of selection bias. These biases are not mentioned in the EPA Issue Paper. Sixth, the pesticide exposures that occur among the AHS cohorts, although most likely intense, are also likely to be intermittent (Blair et al., 2009). However, exposures that occur to the general public, in particular those exposed via contaminated drinking water, would be lower-level but chronic; and exposures to workers employed in the manufacture of triazines would be expected to be both intense and chronic. Finally, in the most recent publication from the AHS, the frequency of assessment and time periods of exposures are not clearly described, and it is unclear if any attempt was made to reflect the relation and nature of time-varying exposure other than using the exposure intensity measures, and no mention of handling time-varying covariates appears. The Panel was informed that lifetime exposure data obtained in Phase I (1993-7) were updated in 2007, which could result in both underestimation and over-estimation of exposure, and covariate data were not updated from Phase I, which, importantly, could have resulted in inadequate control of confounding. The effects of both of these sources of error on effect estimates cannot be accurately predicted.

In addition, in section 3.2.2.2 of the EPA Issue Paper, in the penultimate sentence of the first paragraph, it is not just data collection on family history of prostate cancer that is important because breast and ovarian cancer would have family history of these cancers as important covariates to consider.

On page 39 of the EPA Issue Paper, EPA mentions as a weakness of the AHS that most of the studies were "hypothesis generating." Actually, all of the studies were "hypothesis generating", but all were hypothesis-testing as well.

EPA pointed out correctly that most of the epidemiologic studies of atrazine focused on cancers of the lymphohematopoietic system, the reproductive system, and the endocrine system, and the following comments will assess the evidence from the individual-level studies of these cancer sites.

Prostate cancer was evaluated by the 2003 SAP on atrazine. The Panel's report concluded that the database was insufficient to support a conclusion regarding the potential of atrazine to cause prostate cancer. This conclusion remains valid today even though the AHS cohort studies provide evidence against an association. The reason is that there are still lingering questions about the St. Gabriel triazine manufacturing plant studies. These studies were initiated because of an excess of prostate cancer (5 observed, 2 expected) that occurred prior to the start of the prostate screening program at the plant. The screening program can explain most of the excess cases at the plant, but not the excess represented by these five cases. The exposure experience of these five cases has never been presented. The follow-up case-control study could have compared the exposure experience of these five cases with controls that were employed prior to the screening program, but chose not to do so. As the 2003 SAP report concluded, "Lack of association among farmers does not preclude the existence of a positive association with triazine manufacturing plant workers." This conclusion is based on the likelihood that manufacturing plant exposures are chronic whereas agricultural exposures are more intermittent. Given the possibility that some of the prostate cancer excess cannot be attributed to the plant's screening program, the Panel did not agree with the statement on page 42 of the EPA Issue Paper that there is "consistency of results between manufacturing and agricultural populations" for atrazine and prostate cancer. The Panel believed that, based on the individual-level epidemiologic studies, there still is inadequate information to assess whether atrazine can cause prostate cancer in humans.

Breast cancer and atrazine exposure was evaluated in a study of the AHS cohort and relative risks hovering around 1.0 were observed, providing negative evidence of an association (Engel et al., 2005). A drinking water study in Wisconsin where the mean atrazine level in private wells in the "high use area" was very low, <0.5 parts per billion (ppb), and where only a handful of cases were exposed to well water with ≥3 ppb atrazine, odds ratios (OR) ranged from 1.2 to 1.4 based on small numbers of cases. Although the evidence for an association is extremely weak in this study, it should not be dismissed. Small OR would be expected from a study with very low exposures and considerable exposure misclassification. Moreover, dissimilar findings in the AHS cohort and the Wisconsin study population might be expected, given that drinking water exposures differ from exposures that would occur to pesticide applicators or spouses of applicators. EPA correctly identified the limitations in these studies – small numbers, low statistical power, exposure misclassification, and the inability to evaluate critical time windows of exposure or exposure-response trends. While the EPA Issue Paper does mention the small number of exposed cases in some of the studies as a limitation, it should also be noted that known risk factors for breast cancer and breast cancer subtypes could not be effectively controlled and examined, respectively, because of this important limitation of small

sample size. The latter, consideration of different breast cancer subtypes, is potentially important because breast cancer is not a single disease with only one set of susceptibility factors, and those factors may interact differentially with exposure to change disease risk.

Other limitations not mentioned in the EPA Issue Paper include the predominantly "white" populations that were studied, and the inability to evaluate breast cancer subtypes. The EPA Issue Paper stated that only the AHS study evaluated an a priori hypothesis; but this was true of the Wisconsin study as well.

In summary, the Panel disagreed with EPA's assessment of a lack of evidence for an association between atrazine and breast cancer. The Panel believed that the evidence from the individual-level epidemiologic studies is very weak and it would, therefore, be appropriate to say that inadequate information is available to assess whether atrazine can cause breast cancer.

Ovarian cancer was evaluated by individual-level studies in Italy, the Central Valley, CA, and the AHS cohort. Both the Italian case-control study and the AHS cohort study observed positive associations, and the Italian study observed higher odds ratios (OR) with longer duration of exposure to triazines and a more precise definition of the triazine-exposed group.

The case-control study conducted in an agricultural region of Italy observed an OR of 3.0 based on 7 exposed cases for those "definitely exposed" in the high use region. "Definitely exposed" was assigned to those involved in preparation or use of triazine herbicides or worked in corn cultivation with reported use of herbicides; i.e., it was not "probabilistic" as stated in the EPA Issue Paper (p. 49), but instead was based on questionnaire responses and review by industrial hygienists. Those with more than 10 years exposure had a higher OR (OR=3.1) than those with <10 years exposure (OR=2.1). Those assigned "possible exposure" reported herbicide use or worked in a job involving herbicide exposure, but the herbicide could not be precisely defined, as well as those who denied usage but worked in corn cultivation. ORs were elevated for this group as well but were lower than for the "definitely exposed." Although triazines as a whole were evaluated, reported sales of atrazine were 10 times higher than the other triazine used in the area, simazine. The EPA Issue Paper mentioned the strengths of this study – magnitude of the association, evidence of an exposure-response trend, temporality and biological plausibility – but emphasized the weaknesses, in particular small numbers, the "probabilistic" exposure assessment, and the inability to control for other pesticides. Small numbers would make it harder to detect an association and result in wide confidence intervals, and exposure misclassification would tend to bias results to the null. Adjustment for other pesticides could increase or decrease the effect estimates, but only if the pesticides are independent risk factors for ovarian cancer.

The Central Valley, CA case-control study used Pesticide Usage Reporting data and questionnaire information to construct a job-exposure matrix to assess occupational exposure to atrazine. It also evaluated residential proximity to areas where atrazine was applied. Based on two exposed cases, the odds ratio (OR) for ever occupationally exposed to atrazine was 0.76, and the OR for residential proximity was 0.88 based on 8 exposed cases. The study excluded cases of people who died or were too ill to participate, which EPA correctly pointed out may have introduced a selection bias.

In assessing the epidemiologic evidence for an association between atrazine and ovarian cancer, the EPA Issue Paper states (p. 51) that "Overall, this database presents some indication of a possible association between atrazine and/or triazine use and ovarian cancer; however, the small sample sizes, lack of ability to control for other pesticide use and possibility of unmeasured confounders in the association limit the ability to infer a causal association at this time." The EPA Issue Paper states (p. 68) that "The epidemiologic database for the relation between atrazine and ovarian cancer is small, and weakly suggestive of a possible association across three studies; however, the possible role of random variability, bias or confounding in the risk estimates observed cannot be excluded." The Panel agreed that sample sizes were small in these studies resulting in an inability to control for other risk factors, and wide confidence intervals for the effect estimates. (And of course there is always a "possible role of random variability" even when findings are statistically significant, and in fact, even when exposures are randomly assigned as in experimental studies.) Finally, the Panel agreed that the findings from the better conducted, individual-level studies (i.e., the Italian and AHS studies) are at least "weakly suggestive" of a causal association between atrazine and ovarian cancer. Therefore, the Panel disagreed with the conclusion in the EPA Issue Paper that the "data are insufficient to inform whether an association may exist (p. 69). Given the positive findings in two relatively well-conducted studies, EPA should consider the evidence as suggestive for a causal association between atrazine and ovarian cancer.

With regard to both breast (p. 47) and ovarian cancer (p. 51), it seems inappropriate to indicate lack of confounding bias because the numbers of cases and of exposed cases were probably too small to detect and control for confounding adequately.

Non-Hodgkin's Lymphoma (NHL) and atrazine or triazine exposure has been evaluated in individual-level studies of Midwest U.S. farmers (e.g., Zahm et al., 1993, De Roos et al., 2003, and Schroeder et al., 2001), in the AHS cohort (Rusiecki et al., 2004, Beane Freeman et al., 2011), and in a hospital-based, case-control study conducted in France (Orsi et al., 2009). The 2000 SAP report evaluated the Midwest U.S. studies, including the pooled analysis published in 1993 (Zahm et al., 1993). The 2000 SAP Panel noted the strengths and limitations of these studies and concluded that overall these studies "…suggested that atrazine could produce adverse health effects in humans…."

Since the 2000 SAP review, a re-analysis of pooled data from the Midwest studies has been conducted using standard logistic regression as well as hierarchical regression models to evaluate 47 pesticides, including atrazine, simultaneously (De Roos, 2003). An OR of 1.6 was obtained from the logistic regression model for exposure to atrazine and NHL. A slightly lower OR was obtained from the hierarchical model (OR=1.5). The EPA Issue Paper (p. 55) stated that these findings were "closer to 1.0" than the findings in the earlier pooled analysis, but the reverse is true – the earlier pooled analysis obtained an OR of 1.2 for atrazine use adjusted for 2,4-D and organophosphates (Zahm et al., 1993). The De Roos et al., 2003 study "found only a small number of the pesticides to be risk factors for NHL," which is the likely reason why the study also found that: "Adjustment for multiple pesticides suggested that there were few instances of substantial confounding of pesticide effects by other pesticides." The authors noted findings that suggested superadditive effects of atrazine in combination with carbofuran, diazinon, and

alachlor. In summarizing the study, the EPA Issue Paper stated (p. 55-56): "Overall, DeRoos et al., 2003 cautioned that these data provide only limited evidence of any specific pesticide-NHL associations as results of many different statistical tests, the role of chance may explain these (positive) findings." However, the authors actually stated that the analysis "provided an opportunity to estimate the effects of each specific pesticide and certain pesticide combinations on NHL incidence, adjusted for the use of other pesticides." Moreover, the authors did not state that the evidence was "limited" for the atrazine-NHL association, nor did they state that multiple tests could explain the atrazine findings: "although some of the positive results could be due to chance, the hierarchical regression analysis placed some restriction on the variance of estimates, theoretically decreasing the chances of obtaining false positive results. However, it is possible that the assumptions for the hierarchical regression are too restrictive and that this has increased the number of false negatives." The study was limited by exposure misclassification, likely non-differential, due to the crude exposure metric (exposure/no exposure) and the lack of information on the timing of use of each pesticide. In addition, exposure lags to account for latency could not be evaluated because the timing of pesticide use in relation to disease onset was unknown. Finally, cases and controls were excluded if they had missing data on any of the pesticides evaluated, which could be a source of selection bias, although the authors pointed out that cases and controls were equally likely to be included in the analyses.

Another Midwest U.S. study evaluated the NHL subtypes t(14;18) positive and negative, and obtained an OR equal to 1.7, for the positive subtype based on 15 exposed cases (Schroeder et al., 2001). An OR of 1.0 was observed for the negative subtype. The study had several limitations: (1) it did not evaluate exposure-response trends; (2) it used a crude exposure metric, (atrazine use, yes/no); and (3) over 70% of the cases could not be classified by subtype, requiring a statistical method (Expectation-Maximization) to assign cases to each subtype. Each of these limitations likely would bias the effect estimate towards the null. The study did evaluate several risk factors for NHL to determine if they were confounders, but found that their inclusion in the models did not appreciably change the effect estimates.

The French study observed associations between triazines and NHL (OR=1.9) as well as for the specific sub-groupings, diffuse large cell (OR = 2.1, based on eight exposed cases) and follicular lymphoma (OR=2.3 based on four exposed cases). The study did not report an analysis of exposure-response trend. The findings were not affected by inclusion in the models of several risk factors, including smoking and family history of cancer. Evaluation of different exposure lag periods also did not affect the findings. The EPA Issue Paper (p. 57) and Appendix B (pp. 63-65) emphasized the possibility of selection bias due to the use of hospital controls and the apparent unsuccessful attempt at matching. Unsuccessful matching is not a problem in a case-control study as long as the risk factors used in the matching are included in the analyses, which they were in this study. The authors provide sufficient detail on the comparability of the control series, and provide evidence that it reflected the smoking, alcohol use and the farming practices of the population that gave rise to the cases. Although selection bias cannot be ruled out in any control selection process, it is not a likely source of significant bias in this study. However, exposure misclassification, most likely non-differential, is probably a major source of bias toward the null. Another limitation of this study was that triazines, not specifically atrazine, were evaluated. Finally, regarding any association of atrazine with lymphomas, the text on page

57 should note if any control of confounding factors was included in the analyses of the data from France.

Negative evidence for an association between atrazine and NHL is provided by the recent cancer incidence update of the AHS cohort (Beane Freeman et al., 2011). In a previous evaluation of NHL in this cohort, an association was found between atrazine and NHL, although the exposure-response trend was not monotonic, most likely due to exposure misclassification (Rusiecki et al., 2004). With updated cancer incidence data, this association disappeared, with RR < 1.0 in the upper quartiles of exposure (Beane Freeman et al., 2011). Moreover, no associations were found with NHL subgroupings.

Although evidence from the AHS cohort does not suggest a causal association between atrazine and NHL, the studies conducted in the Midwest U.S. and in France provide positive evidence. Given this positive evidence, the Panel disagreed with the EPA Issue Paper (p. 59) that "overall, the database lacks evidence of an association" between atrazine or triazine exposure and NHL or its sub-types. Instead, the Panel concluded that there is suggestive evidence for a causal association between atrazine and non-Hodgkin's lymphoma.

For other lymphohematopoietic cancers, the evidence is negative for leukemias (except hairy-cell leukemia) and multiple myeloma.

Hairy-cell leukemia and triazine exposure were evaluated by two hospital-based, case-control studies conducted in France (Clavel et al., 1996; Orsi et al., 2009). Both studies used similar methods for control selection and exposure assessment. In the earlier French study (Clavel et al., 1996), an odds ratio (OR) of 2.4 was observed for "definitely exposed" to triazines based on 20 cases. When the analysis was restricted to cases and controls unexposed to organophosphates, the OR ranged between 1.5 and 2.0. No exposure-response was observed, but the data were not presented. The more recent French study (Orsi et al., 2009) observed an OR of 5.1 for hairy-cell leukemia based on four triazine-exposed cases. As stated previously, this study evaluated several exposure lag periods, but did not report an analysis of exposure-response. The EPA Issue Paper emphasized the problem of systematic bias in the use of hospital controls, but both studies adequately described their control selection process and provide evidence against selection bias. Although systematic bias in the control-selection process can never be entirely ruled out in any case-control study, it does not appear to be a source of major bias in these studies. Instead, the major source of bias in these studies is non-differential exposure misclassification. Other limitations include small numbers of exposed cases resulting in wide confidence intervals, an inability to adequately assess exposure-response trends, and limits on the ability to adjust for confounders in the regression models. The Panel disagreed with the EPA Issue Paper assessment that there is a lack of evidence of an association between triazines exposure and hairy-cell leukemia. The Panel concluded that these two studies provide suggestive evidence for a causal association between triazines and hairy-cell leukemia.

Thyroid cancer and atrazine exposure was evaluated in the recent cancer incidence update of the AHS cohort (Beane Freeman et al., 2011). This is the first study to evaluate this association. Elevated RRs were observed within the highest 3 quartiles of exposure lifetime days, but the exposure-response trend was not monotonic, most likely due to non-differential

exposure misclassification bias. A non-monotonic trend was also observed for intensity-weighted lifetime days. Unfortunately, the categorization of exposure used in the analysis of thyroid cancer was based on all cancer cases in the cohort, and was not appropriate for the analysis of thyroid cancer. The Panel recommended that the researchers re-analyze these data using more appropriate categorical cutpoints (e.g., using smoothing methods such as splines or LOESS to evaluate the shape of the exposure-response curve). The Panel concluded that this study provides suggestive evidence for a causal association that needs to be followed up.

Brain cancer/gliomas and atrazine exposure was evaluated in a few individual-level studies, and the evidence was negative for a causal association (Rosenman and Talaska, 2005). Other cancers (e.g., lung, pancreas, melanoma, and colorectal) and atrazine exposure have only been evaluated in the AHS cohort, and the evidence so far is negative. However, in the recent cancer incidence update for the AHS cohort (Beane Freeman et al., 2011), positive but non-monotonic exposure-response trends were found for liver cancer and esophageal cancer, but the evidence is inadequate to assess whether atrazine can cause these cancers and requires follow-up.

Childhood cancers and paternal use of atrazine were evaluated in the AHS cohort (Flower et al., 2004). A standardized incidence ratio (SIR) of 1.27 was observed for all childhood cancers combined. Childhood cancers were not evaluated separately, and the exposure metric was crude (yes/no paternal exposure). Maternal use of atrazine could not be evaluated. A study in California evaluated residential proximity to areas where triazines were applied and occurrence of childhood acute lymphocytic leukemia; a positive but non-monotonic exposure-response trend was observed when lifetime exposure (over the child's lifetime) was evaluated, but any excess, when exposure over the first year of life was evaluated, was not observed (Rull et al., 2009). Both studies had important limitations; in particular, the likelihood of considerable exposure misclassification. The Panel concluded that these studies provide inadequate evidence to assess whether atrazine can cause childhood cancers.

In summary, the Panel concluded that the cancers for which there is suggestive evidence of carcinogenic potential include: ovarian cancer, non-Hodgkin's lymphoma, hairy-cell leukemia and thyroid cancer. These cancer sites require follow-up studies. In addition, cancers for which there is inadequate evidence include: prostate cancer, breast cancer, liver cancer, esophageal cancers, and childhood cancers. These cancer sites also require follow-up studies. Cancers not likely to be caused by atrazine given the currently available data include: oral, lung, colorectal, pancreas, bladder, leukemias (except hairy-cell leukemia), multiple myeloma, melanoma, kidney, larynx, and brain/gliomas.

Charge Question 11 - There is no compelling evidence in the available experimental toxicology database that atrazine may be carcinogenic in humans. This database was reviewed by the SAP in April 2010. Briefly, the results of both guideline and non-guideline in vivo as well as in vitro studies do not suggest atrazine has mutagenic or carcinogenic properties. With regard to mammary tumors, several internal and external peer reviews have concluded that the mode of action through which atrazine influences mammary tumor development is not operational in humans. Further, EPA's current review of the atrazine cancer epidemiology database did not identify evidence across any of the cancer-specific databases evaluated as to the carcinogenic potential of atrazine that EPA finds

sufficiently convincing to change its conclusions. Therefore, in view of the evidence in the experimental toxicology and epidemiologic databases, EPA concludes atrazine is not likely to be carcinogenic in the human population. The observational data lend further support for the human relevance of the laboratory rodent tumor findings, i.e., the databases lack evidence of an association between atrazine and cancer in the human population.

a) Please comment on the extent to which the scientific information supports the integrative analysis contained in Section 3.3 with respect to the similarities, differences of the experimental toxicology and epidemiologic findings. Please comment on any significant uncertainties in the epidemiologic findings.

Panel Response:

The integrative analysis of the experimental and epidemiologic studies and their results that is contained in Section 3.3 of the EPA Issue Paper reflects well the state of the available scientific information regarding the relation of atrazine exposure to cancer risk. The known neuroendocrine health effects of atrazine provided the rationale for focusing on prostate, breast, ovarian and thyroid cancer, but the relation of atrazine exposure to other cancers, including lymphohematopoietic malignancies, was also evaluated.

The Panel raised concerns about integration when the animal mechanistic or toxicological data are lacking, and only epidemiologic data are available (e.g., with thyroid cancer), or if the animal mechanistic or toxicological data are inconsistent or reflect different mechanisms from those in human epidemiologic data (e.g., with breast and ovarian cancer). The February 2010 SAP attempted to provide a framework for integration of these data, based largely on the Bradford Hill criteria, and recommended that a weight of evidence evaluation be used to build a case for each health outcome using that framework. The approach outlined in this framework was used to prepare the epidemiology assessment for atrazine presented to this Panel. The framework is presented in Figure 1 below, recreated from the framework document, but with the addition of the "INFORM line" which has been added for this discussion. OPP's draft framework "proposes to use the Bradford Hill Criteria as modified in the Mode of Action (MOA) Framework as an organizational tool for describing and reviewing data from animals and humans" and "taking into account factors such as dose-response and temporal concordance, biological plausibility, coherence and consistency (EPA, 2010)." Using this analytic approach, epidemiologic and human incident findings can be evaluated in the context of other human information and experimental studies to evaluate consistency, reproducibility, and biological plausibility of reported outcomes and to identify areas of uncertainty and future research. Biological plausibility, i.e., animal mechanistic or toxicological evidence, is only one component of many in that framework, and the other components (strength of the association, temporality, exposure-response relation, consistency of results) must also weigh in heavily. Thus, if only epidemiologic data are available, these latter considerations should be applied and weighed heavily even if biological plausibility or animal experimental data are not available, i.e., the epidemiologic data should not be ignored in the absence of mechanistic or toxicological evidence. Numerous examples can be found in the public health literature of making preventive policy in such situations. For example, the situation with smoking and lung cancer 50 years ago

was a case of implementing preventive actions in the absence of sufficient pathophysiologic mechanistic data.

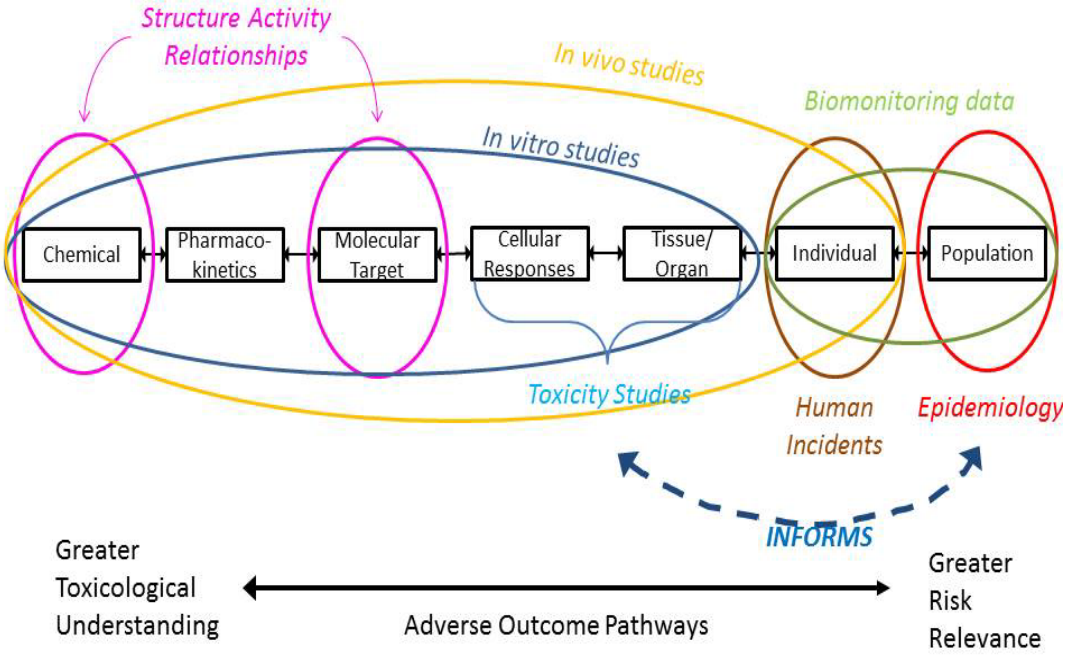


Figure 1. Source to Outcome Pathway: Chemical effects across levels of biological organization (adopted from NRC, 2007 and presented at February 2009 SAP meeting).

As the Panel struggled with integrating the findings from the assessment of the human epidemiology studies, there was a realization that very little had been provided in the way of hypothesized adverse outcome pathways leading to cancer outcomes. Pathways leading to breast cancer and prostate cancer in animals are known. The human epidemiology research suggests other cancer targets, such as the thyroid, and AOPs for these cancer outcomes need to be hypothesized and related in vitro and in vivo research assessed.

Many on the Panel believed that the epidemiology data failed to provide compelling evidence that atrazine is not carcinogenic. While *in vitro* and animal *in vivo* studies can inform epidemiology, human epidemiology findings can inform the need for additional *in vitro* or *in vivo* studies to explore hypothesized modes of action that were possibly not considered before the full epidemiology assessment was completed. In this way, the human epidemiology informs which molecular target, cellular and tissue/organ study data needs reviewing. Weak epidemiology studies with findings that suggest a strong relationship between cancer and atrazine (as is the case with the one thyroid study) should lead to additional human epidemiology research or detailed animal studies and should not be ignored because it may not fit the currently hypothesized mode of action derived from animal experiments in which the mode of action may be different than in humans. The whole point of the framework exercise is to identify gaps where lack of data may lead one to miss important health effects. It was pointed out, for example, that most of the epidemiology data were from studies of occupationally exposed

populations, primarily middle-aged, white working males, and thus may not be applicable to women or other racial/ethnic groups.

A second issue that was discussed was that the EPA Issue Paper focused almost exclusively on a single mechanism of action, the neuroendocrine pathway and suppression of the LH surge, which largely has relevance only to some reproductive outcomes and a few other potential outcomes. Indeed, the MOA section of the EPA Issue Paper emphasizes reproductive, rather than cancer effects. It was recommended that a wider search for potential mechanisms for such potential cancer mechanisms as those for thyroid cancer and hairy-cell leukemia be undertaken. This MOA may have relevance for some cancer targets (such as female breast and prostate), but not all of the cancer targets. Where are the other cancer relevant MOAs and what is the evidence that has excluded them from consideration? Again, many on the Panel believed that if the epidemiology was being used to exclude consideration of other MOAs, then this decision might not be supported by the epidemiology assessment.

A brief summary and review of integration of results by cancer type to support this conclusion follows. The accompanying table shown in Appendix A provides a summary of the studies and their findings and is the format the Panel recommended OPP use for presentation of these findings.

For prostate cancer since the EPA review in 2003, which reported a positive association of atrazine exposure in an occupational cohort that was largely attributed to enhanced prostate cancer screening (MacLennan, 2002), two nested case-control studies have been reported, one in a manufacturing cohort (Hessel et al., 2004) and one in the AHS (Alavanja et al., 2003). Neither altered the conclusion from the 2003 review. One ecologic study (Mills, 1998; Mills and Yang, 2003) suggested an association with atrazine or triazines, but this design is considered the weakest, particularly from an exposure assessment perspective, and thus is appropriately down-weighted in the weight of the evidence assessment. Two additional publications from the AHS (Rusiecki et al., 2004; Beane Freeman et al., 2011) confirmed the previous findings of no association and provided no evidence of an exposure-response trend. The EPA Issue Paper correctly concludes that these results, combined with the lack of a prostate cancer effect in animal chronic toxicology studies and lack of experimental evidence about the biological plausibility of atrazine being carcinogenic to the prostate, largely argue against an association of atrazine exposure with prostate cancer risk. However, as noted in the limitations above, these study samples have comprised largely white males and thus did not include the subgroup at highest risk for prostate cancer (African Americans), so that no conclusions can be made about generalizing the results published to date to such high risk groups. In addition, while the prostate screening may explain some of the excess of cancers in the manufacturing cohort, one cannot be certain that it explains all of the excess.

For breast cancer, a rationale once existed for exploring the possibility of a relation in the known neuroendocrine health effects of atrazine and results from experimental studies showing effects on mammary gland tumors in female rats and on mammary gland development, but the mode of action in rats has subsequently been shown not to apply in humans. Since the EPA review in 2003, two ecologic studies (Muir et al., 2004; Mills and Yang, 2006) (in addition to two ecologic studies published before and considered in 2003), one population-based case-

control study (McElroy et al., 2007) and two assessments of the AHS data (Engel et al., 2005; Beane Freeman et al., 2011), have been published that have assessed breast cancer risk in relation to atrazine exposure. The ecologic studies had inconsistent findings, with some indicating a small but significant elevation in risk, and others indicating no association of atrazine exposure with breast cancer risk. Given the weaknesses of the ecologic design, these studies carry little weight in the totality of the evidence, and their inconsistent results underscore the important possibility of misclassification of exposures. The population-based case-control study was conducted in a sample of Wisconsin women and found no significant elevation in breast cancer risk associated with atrazine exposure from well water, although the design provided possibilities for exposure and outcome misclassification. A nested case-control analysis of female spouses of pesticide applicators in the AHS and a cohort analysis of the AHS that included female pesticide applicators who could have direct exposure both showed no significant relation of atrazine exposure to breast cancer and no evidence of an exposure-response trend, although the number of exposed breast cancer cases was too small to permit a meaningful evaluation of this or of subtypes of breast cancer for differential susceptibility. Further, in general, data on chronic, low level atrazine exposure in individuals, e.g., via drinking water, in relation to health outcomes are largely lacking. Thus, while a few studies did show very weak evidence, the available literature largely does not support an association of atrazine exposure with breast cancer risk; however, the weak evidence should not be dismissed.

For ovarian cancer, the 2003 EPA review reported on an earlier case-control study conducted in Italy which demonstrated a greater than two-fold significantly elevated odds of histologically confirmed ovarian cancer in women classified as definitely exposed to triazines with evidence of an exposure-response relation with duration and probability of exposure (Donna et al, 1984 and 1989). Two ecologic analyses published prior to 2003 reported no increased risk of ovarian cancer associated with atrazine or triazine use (Hopenhayn-Rich, 2002; Van Leeuwen et al., 1999). A case-control study conducted in California using indirect measures of triazine exposure also found no significant association, although the numbers of cases and controls assessed to be exposed was too small to have sufficient statistical power to detect an association if it truly existed (Young et al., 2005). The recently published cohort analysis from the AHS indicated a non-significant nearly three-fold increase in the odds of ovarian cancer in female applicators, but the number of cases was too small to detect this as statistically significant, to detect and control for confounding adequately, or to examine exposure-response relationships (Beane Freeman et al., 2011). Thus, while the existing literature suggests a possible association of atrazine and/or triazine with the rare ovarian cancer, the methodologic concerns regarding these studies provide only a suggestion of an association that requires confirmation in larger, better designed studies.

For thyroid cancer, only one recent cohort analysis from the AHS (Beane Freeman et al., 2011) provided information on a possible association with atrazine use and only in male applicators because, even though women are at greater risk of thyroid cancer, only one female applicator in the cohort was exposed to atrazine. The results of these analyses were highly suggestive of a positive relation between atrazine exposure and thyroid cancer with a greater than four-fold increased odds of the disease after controlling for multiple potential confounding factors. An effect of this magnitude is unlikely to result from inadequate control of confounding. An exposure-response trend was observed that was non-significant and non-monotonic, probably

due to the small numbers of this relatively rare cancer but also possibly due to using exposure cutoffs for all cancers. A better approach might have been either to collapse some exposure categories to have more cases per category or to use tertiles of exposure or examine risk estimates for thyroid cancer above and below the median exposure level. A biological mechanism is not known for this relation, but this is a relatively minor limitation, given that relatively little is known about the etiology of thyroid cancer in general with the exception of the association with radiation exposure and with family history. Given this very suggestive finding from a single study, while not sufficient to be certain of a causal relation, it would be inappropriate to say that atrazine is unlikely to be carcinogenic for the thyroid. Thus, these findings require replication in a larger study and more experimental investigation with regard to potential biological mechanisms.

For lymphohematopoietic cancers, the 2003 EPA review included results from several population-based case-control and cohort studies of the relation of atrazine or triazine exposure to non-Hodgkin's lymphoma (NHL) and largely found no association. Since that review, a pooled analysis of case-control studies (DeRoos et al., 2003), two hospital-based case-control studies (Clavel et al., 1996; Orsi et al., 2009) and one population-based case-control study (Schroeder et al., 2001) have been published about the relation of pesticides, including atrazine, to NHL and to leukemias and lymphomas, as well as two cohort analyses about leukemias and lymphomas from the AHS. The earlier population-based case-control studies found no association of atrazine or triazine use with leukemia or multiple myeloma (Brown et al., 1990 and 1993), but both of these studies and the pooled analysis did find a significant 1.5-fold to 2.5-fold elevation in odds of histologically-confirmed NHL with triazine use after adjusting for other pesticide exposures (Cantor et al., 1992; Hoar et al., 1986), as did a population-based, case-control study (Schroeder et al., 2001), although the numbers of exposed cases in most of these studies was small. One of the earlier hospital-based case-control studies examined the relation of pesticides, including triazines, to hairy cell leukemia (HCL) and found a 2.4-fold significantly increased odds with ever use of triazines, which was attenuated to a non-significant two-fold increase when the analysis was restricted to never users of organophosphates (Clavel et al., 1996). The more recent French case-control study (Orsi et al., 2009) found elevations in odds of HCL, NHL, and follicular lymphoma, but only the five-fold elevation in odds for HCL was statistically significant, and no association was found with chronic lymphocytic leukemia, although the numbers of cases for each of these malignancies was fairly small. Non-significantly elevated adjusted odds ratios of about 2.0 were also found in earlier cohort analyses of the AHS (Rusiecki et al., 2004) for NHL and multiple myeloma with a positive but non-significant exposure-response trend, but these were not duplicated in the most recent cohort analyses from the AHS which had twice as many cancer cases (Beane Freeman et al., 2011). Thus, while early studies suggested possible relationships of atrazine use with NHL and HCL, the more recent, better designed and controlled studies with larger sample sizes did not replicate these findings, indicating that evidence for associations of atrazine with NHL and HCL is limited in humans or in animal experimental studies.

A few other cancers have been investigated for possible associations with atrazine and/or triazine use. These investigations have included a relatively large case-control study of adult gliomas that found no significant association in women (Carreon et al., 2005) or men (Ruder et al., 2004) and two ecologic studies that found a moderate relation and no association (Mills,

1998; Van Leeuwen et al., 1999) so that the weight of the evidence suggests no association. Additionally, ecologic studies have demonstrated moderately elevated non-significant associations of pesticides or atrazine use in relation to pediatric bone cancer and leukemia (Thorpe and Shirmohammadi, 2005). A population-based case-control study from northern California found a nearly two-fold significantly increased odds of pediatric acute lymphocytic leukemia associated with moderate triazine exposure, which increased with adjustment for exposure to other pesticides, but no association with the high exposure category (Rull et al., 2009). Earlier analyses of the AHS were not cancer site-specific but found a significant modest increase in risk (odds ratio=1.36) for all cancers and significant two-fold increase in risk for lymphomas (Flower et al., 2004). Earlier ecologic (Van Leeuwen et al., 1999), case-control (Hoar, 1985) and cohort (MacLennan et al., 2002 and 2003) studies have found no significant association of atrazine with colon or rectal cancer, and this was confirmed in a more recent, large case-control study (Lee et al., 2007) and cohort analysis in the AHS (Beane Freeman et al., 2011).

Associations of atrazine exposure with other cancer sites, including oral, esophageal, pancreatic, melanoma, renal, laryngeal, brain, lung, bladder and liver cancer have been investigated recently in the AHS, and no significant associations were found (Beane Freeman et al., 2011). Thus, the EPA Issue Paper correctly concludes that the evidence is currently insufficient to determine if atrazine exposure increases the risk of pediatric cancers, and the current evidence supports no association with adult gliomas, colon or rectal cancer, or oral, esophagus, pancreas, melanoma, kidney, larynx, brain, lung, bladder and liver cancer. However, one Panel member asked, based on atrazine stimulation of the adrenal gland, whether adrenal cancers were surveyed.

b) Please comment on whether the epidemiology literature published since the last SAP review including the AHS findings is sufficient to justify changing the Agency's conclusions that atrazine is not likely to be carcinogenic to humans.

Panel Response:

The epidemiologic evidence compiled since the last SAP review regarding the carcinogenicity of atrazine does not justify changing the Agency's conclusions regarding prostate cancer, breast cancer, adult gliomas, oral, esophageal, pancreatic, melanoma, renal, laryngeal, brain, lung, bladder, colon, rectal, and liver cancer or leukemia, chronic lymphocytic leukemia or multiple myeloma.

The epidemiologic evidence regarding a potential association of atrazine exposure with ovarian cancer is suggestive, but still inconclusive and requires more rigorous investigation with larger sample sizes, which is very difficult for this rare cancer that is likely to have a long latent period, which also greatly complicates exposure assessment.

For thyroid cancer, only one study, the recent AHS cohort analysis, is available, but it suggests a strong relationship (fourfold increased adjusted odds of thyroid cancer), that is unlikely to be due to residual confounding and a non-significant exposure-response relationship. This strong and suggestive finding from a single study is not sufficient to be certain of a causal

relation between atrazine and thyroid cancer and thus requires replication in a larger study and more experimental investigation with regard to potential biological mechanisms.

The epidemiologic findings regarding an association of NHL and HCL with triazine use after adjusting for other pesticide exposures, although having small numbers of exposed cases in most studies, suggest about a 1.5- to two-fold increase. Some of these estimates were statistically significant, and some findings had non-significant exposure-response relationships, although the numbers of cases for each of these malignancies was fairly small. However, these findings were not duplicated in the most recent cohort analyses from the AHS which had twice as many cancer cases (Beane Freeman et al., 2011). Thus, while early studies suggested possible relationships of atrazine use with NHL and HCL, the more recent, better designed and controlled studies with larger sample sizes did not replicate these findings, indicating, as mentioned in the EPA Issue Paper, that sufficient evidence for associations of atrazine with NHL and HCL is limited in humans or animal experimental studies, although the limitations of the AHS noted in response to Question 10 should not be ignored.

Ecologic and case-control epidemiologic studies of pediatric cancers have had suggestive findings for an association of atrazine exposure with increased risk of acute lymphocytic leukemia in children, although a monotonic exposure-response relationship was not observed. Thus, the EPA Issue Paper correctly concludes that the evidence is currently insufficient to determine if atrazine exposure increases the risk of pediatric cancers, particularly leukemia.

The Panel recommended adjusting the conclusion that atrazine is unlikely to be carcinogenic to humans to "inadequate information to assess carcinogenic potential." This category is appropriate given that an association is unlikely (or possibly very weak due to inadequate information) with some cancers, as indicated above, but epidemiologic evidence suggests possible associations with NHL as well as ovarian and thyroid cancers. The Panel noted that, if possible, it would be useful and appropriate to make conclusions for individual cancers as opposed to making a blanket determination for cancer in general. Given the strong association with thyroid cancer in the epidemiologic study, a categorization of "suggestive evidence of carcinogenic potential" would be appropriate until other studies are conducted. In addition, given the suggestive association with ovarian cancer in the AHS, it would also be appropriate to use the category "suggestive evidence of carcinogenic potential". Although the Young et al., 2005 study did not find evidence of an association with ovarian cancer, the study was limited by small numbers of exposed women and non-differential misclassification bias which would bias results toward the null.

Additional concluding remarks were noted during the SAP's discussion of Charge Question 11 and related to issues of previous charge questions. With the February 2010 SAP, the EPA showed commendable evidence of initiating a cultural change in developing a framework for including epidemiologic study results in its risk assessment. The Agency is to be applauded for this effort because such data reflect risk situations in human populations, which is particularly important when animal mechanistic and toxicological data do not exist or are inconsistent or reflect mechanisms or pathways that do not apply to humans. However, in general, cultural change is a lengthy process because it follows decades of thinking about and approaching risk assessment one particular way and the Agency does not appear to have fully embraced this change. Specifically, the lack or inconsistency of animal mechanistic and

toxicological data with results from human epidemiologic data does not mean the risk associations identified in human studies do not reflect reality, even though animal experiments are not available or do not support the epidemiologic findings because animal models do not always apply to humans even when they are available. Furthermore, notable epidemiologic findings (using the framework established in February 2010) should be given greater weight in risk assessments and should suggest avenues for future mechanistic and toxicological investigations if these are lacking, as is often the case. In addition, the Panel noted that this seems achievable, and particularly important in light of the uncertainties in existing epidemiologic evidence, to significantly expand epidemiologic information by evaluating existing information on atrazine in public water supplies, populations served, and patterns in health outcomes. This approach is especially relevant for atrazine because drinking water is a primary avenue of exposure. The Agency may make significant inroads by conducting such a study in collaboration with epidemiologic researchers.

Charge Question 12 - The proposed refined dose response assessment for atrazine will be based on internal measures of exposure. At the September 2010 SAP meeting, the Agency presented estimates of area under the concentration-time curve (AUC) for plasma triazines based on a rat [14C]-atrazine study (i.e., Thede 1987). The Panel concluded 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." The Panel also noted that "the use of total chlorotriazine based on total 14C-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 response to the comments by the Panel, the Agency has evaluated additional pharmacokinetic studies involving different species including humans and two additional rat14C-atrazine studies that support the plasma clearance estimates obtained from the Thede 1987 study. Based on the consistent linear pharmacokinetic behavior of plasma triazines resulting from orally administered doses of atrazine across the different studies, the Agency proposes to use an interim pharmacokinetic modeling approach based on a one-compartment linear model to inform internal dosimetry that seems to adequately describe the pharmacokinetics of plasma triazines in rats from orally administered atrazine.

a) Please comment on the strengths and limitations associated with this simplified pharmacokinetic modeling approach for human extrapolation.

Panel Response:

Simplicity of the model is, in itself, a strength. Generally, the simplest model (i.e., one having the fewest pharmacokinetic parameters) that fulfills the purpose for which the model is used is preferred over more complicated models. This approach has minimal data requirements and some of the studies considered by the Agency provide data sufficient to support only a simple model. The available data have utility where the purpose of the model is to predict human exposure to atrazine and its metabolites on the basis of information obtained from rats exposed to atrazine via known dosing regimens. The one-compartment model is adequate for

this purpose, given that area under the plasma concentration time curve (AUC) provides an appropriate measure of internal exposure. Generally, the available data conform to one-compartment model behavior, and it is possible to obtain model parameter values in rat, monkey, and to a limited extent, in human.

While the simplicity of the one-compartment model is a virtue, it tends to obscure some details of the pharmacokinetics. There are some areas of uncertainty since the concentration(s) of atrazine and its metabolites at sites of action (where the chemical initiating event occurs) are not known. Nor is it known whether saturable binding, transport, and metabolism occur, so as to make the exposure of sites of action depend nonlinearly on atrazine dose, duration of exposure and route of administration. If they are discovered to be important to the hazard assessment of atrazine, the one-compartment model may not be adequate to characterize their influence. The fact that the half-life of total radioactivity is dose-independent over a broad range of doses that encompasses environmentally relevant exposures provides some confidence that such nonlinearities probably do not exist.

As discussed below in the response to the next Charge Question, it is currently difficult to identify a clear link between pharmacokinetic behavior and pharmacodynamic activity for suppression of the LH surge (one of the secondary lesions resulting from an unidentified primary lesion). Under these circumstances, it is reasonable to examine available pharmacokinetic data in the way proposed by the Agency. AUC is the dose metric that represents the exposure of all tissues (target and non-target) to the toxicant. This approach has advantages of simplicity since it pools all of the concentrations of parent compound and metabolites in the plasma to provide one measure of exposure at target sites and requires no assumptions about the location of the site of action. Further, where radiolabel is used, it is possible to check for mass balance. This provides some reassurance that the entire administered dose is accounted for.

In general, the Panel supported the work of the Agency in pursuing a dose-response analysis based on an internal dose metric, as an alternative to administered dose in the interest of reducing uncertainty in inter-route, interspecies and intraspecies extrapolations. In the one compartment model, the Agency has attempted to maximize use of the data available to them, making mainly conservative choices in the absence of hard information and verifying estimates to the extent possible. However, it was noted that the one compartment model was proposed by EPA as an interim approach, given that a thorough review and evaluation of a recently submitted PBPK model by Syngenta has not been completed. It has also been noted by the Agency that a verified PBPK model would constitute the “ideal approach”.

b) Compare and contrast the strengths and weaknesses of using total radioactivity for a pharmacokinetic analysis (as presented in the Agency’s Issue Paper) as opposed to using available pharmacokinetic data for the parent and the chloro-s-triazine metabolites that have similar toxicological properties to the parent?

Panel Response:

Generally in pharmacokinetic studies, there are difficult choices to make because the two commonly adopted approaches, PBPK modeling and PK modeling with fewer compartments,

offer different advantages and disadvantages. It is often necessary to opt for a middle path between a complete physiologically based system, and one or two compartment models. Complex (arguably, more realistic) PBPK models require significant amounts of work to parameterize them. Sample preparation involves dissection of individual body compartments, quantitative extraction, and preliminary clean-up stages before analysis. It is advisable to use labeled analytes to correct for matrix effects in the mass spectrometric analysis. Even so, it is difficult to achieve a mass balance. Modern sensitive analytical techniques using powerful mass spectrometers as detectors can achieve low levels of quantification, and allow the analysis of both parent compound and metabolites in one suite. However, much lower levels of quantification are possible by using radiolabeled compound with combustion and scintillation counting. The strengths of using total radioactivity lie in the knowledge that no toxicologically active atrazine metabolite is missed and in the sensitivity and ease of its quantification, which permits comprehensive characterization of AUC and exposure of the animal even at low doses.

Concentrations of analytes measured using chromatographic techniques and MS are normally those of free (unbound) material that is available for distribution, redistribution and elimination. This is the toxicologically relevant fraction in terms of exposure of a site of action. Assuming a mammilary model with the circulatory system providing a rapid mass transport of material to and from all tissues, then when a steady state is achieved, the levels in all of the tissues will change at matching (not necessarily the same) rates. The area under the plasma concentration curve represents the overall opportunity for exposure of the site of action whether located in one or more tissues. The PBPK model thus provides information which is readily interpretable and provides support for interpreting mode of action studies. It does not have to assume that parent compound and metabolites are equitoxic. It can also help to identify where deviations from expected behavior are caused due to changes in the physiology (e.g., changes in renal function, changes in blood flow) caused by toxicological lesions

Simpler models based on fewer compartments require fewer samples to be analyzed. Where radiolabel is used, less sample preparation is required. Measurement of total radiolabel has benefits since sample preparation is easy, and scintillation counting is sensitive and can be automated. Mass balance can be readily checked. However, there are some areas where clear interpretation is not possible.

The discrepancy in plasma clearance between the sum of metabolites measured by MS and total ^{14}C may be methodologically determined. That is, different fractions (free, and the sum of free and bound, respectively) are measured by the two methods One way to determine whether there is a significant bound fraction would be to ultrafiltrate the plasma and to count the filtrate and the filter separately. If total radioactivity AUC calculations had excluded bound radioactivity, it is likely that the total radioactivity and total chloro-s-triazine AUCs would be similar. The radiolabel approach could have advantages over limiting the exposure assessment to the plasma concentrations of atrazine and individual chloro-s-triazine metabolites since there are clearly additional atrazine-derived metabolites in plasma that are not included among the chloro-s-triazines measured in the PBPK modeling study. Further, the half-lives of these additional metabolites are considerably longer than those of the chloro-s-triazines. However, if the “additional metabolites” are plasma protein adducts, the measured fraction will include toxicologically relevant material and material that is not available for distribution to the site of

action, and this advantage of total radioactivity would disappear. If there is significant binding then, the exposure of the site of action would be overestimated by the total radioactivity method. If binding sites become saturated, the AUC will represent the biologically available material plus a constant as the exposure metric. If the "additional metabolites" included additional toxicologically active metabolites, using only the chloro-s-triazines would tend to underestimate the exposures of concern to atrazine and its metabolites. Under those conditions, total radioactivity would be a more comprehensive metric for exposure. Overall, while the modeling is simpler using total radiolabel, interpretation in terms of toxicology and mode of action is more difficult.

Some other factors need to be considered carefully when using some of the old ^{14}C-labeled atrazine studies. It is reasonable to assume first order kinetics for the overall elimination process, and there is good evidence for this from the studies presented by the Agency and by Syngenta. However, there is evidence in some of the radiolabel studies that two first order elimination processes are proceding simultaneously. A double exponential model would, therefore, be expected to fit the elimination data, with a small fraction (bound material) being operated on by a lower fractional rate constant, and a larger fraction (free material) being operated on by a higher rate constant. The former may possibly represent elimination of the fraction of the major analyte (DACT) that is bound to proteins, and the latter elimination of the free material dissolved in the plasma. There is evidence from the radiolabel study in a non-human primate presented by the registrant for the double exponential nature of the elimination process. There is also evidence from the rodent radiolabel data used by the Agency to estimate the fractional elimination rate constants. In the Ln (concentration) linear time plots, the pattern of residuals from the fitted straight lines is consistent with double exponential behavior. This is also seen for DACT in the mouse study of Ross et al., 2009. However, there are insufficient points (degrees of freedom) to fit anything other than a single first order model. Fortunately, the low rate constant operates on only a small fraction of the compound, and so the bias produced in the estimate of the elimination rate constant used by the Agency will not markedly affect the interpretation of the data. The analysis of the available radiolabel studies to examine the consistency of the estimates (all based on a very few points) of the elimination rate constant was a sound approach, and this is evidenced by the close agreement between the estimates presented in the EPA Issue Paper. The only deviations are the high elimination constants observed at the 50 mg/ kg/day and 100 mg/ kg/day doses in the Thede (1987) data.

A weakness of modeling total radioactivity is that the label is distributed among multiple chemical species, each having its own pharmacokinetic behavior. It is unlikely, however, that all species, or even that two of them, have similar pharmacokinetic behavior. The dominant species at a particular time will dominate the kinetic behavior of the total radioactivity, and so total radioactivity tends to "hide" much of the underlying kinetic behaviors. The approach taken by the Agency assumes that the area under the plasma concentration curve reflects the opportunity for exposure of the site of action of atrazine and some of its metabolites. It also assumes that the toxicities of the metabolites are similar. From what is known of the LH-surge suppression activities of atrazine and some of the metabolites, they appear to be approximately equipotent. It is reasonable, therefore, to use the AUC of total radioactivity as the exposure metric. The one-compartment model well characterizes the plasma concentration-time data of total radioactivity after ^{14}C-atrazine administration, and the concentrations of atrazine, DEA, DIA, DACT and their

sum after oral gavage and dietary administration of atrazine. Further, model parameter (ke, Vd) values in rat and monkey seem to be independent of the size of the dose over a broad range (1-100 mg/kg, pp. 89). While not ideal, the simple model seems to provide a reasonable approach to use with the information currently available, particularly as one metabolite (DACT) dominates the profile.

Caution needs to be used when interpreting the available radiolabel studies based on equally spaced constant doses and small numbers of infrequent samples. These give a very poor definition of the pharmacokinetic profile. The plateau produced by joining the points at 24 h intervals cannot be anything but a smooth line. However, in studies with more frequent sampling (as those undertaken by Syngenta), the profile comprises a series of sharp pulses where dosing is by oral gavage. Despite their shortcomings, these studies can be useful since the area under the apparent plateau seems to form a consistent proportion of the total AUC.

The Syngenta PBPK data show that when atrazine was administered by oral gavage the AUC was larger than that found for dietary dosing, particularly for the parent compound and the monodealkylated metabolites. However, the major differences were that a smoother plasma concentration curve with only small fluctuations following feeding was produced by the latter technique, and much higher transient peaks by gavage. Given the relatively modest differences in overall AUC, it was surprising that a suppression of the LH surge was produced by gavage administration but not by dietary administration. Several possible explanations for these observations include the following:

- Total AUC may not be the appropriate measure of exposure; it may be the area under the curve above a critical threshold concentration
- A sustained constant low concentration may not be sufficient; it may require short pulses of much higher concentration
- The dietary route takes 24 hr to reach the high plateau concentration, and this then continues (because of the feeding pattern) after the concentration from the last dose by gavage has fallen markedly (this may be a factor where critical periods of exposure are involved).

The significant difference between the kinetic behavior of total radioactivity and the sum of atrazine, DEA, DIA, and DACT indicates that there may be more to discover about the atrazine species that comprise the total radioactivity and their pharmacokinetic behavior. In light of the data presented by Syngenta comparing dosing by oral gavage and by the dietary route, it would be worth re-examining the assumption that AUC as measured by chromatography-mass spectrometry or total ^{14}C-label is an appropriate measure of the exposure required to suppress the LH surge. Given the importance of identifying an appropriate dose metric, further work to explore these results would be justified. It may be possible to distinguish between some of the options presented above by using different routes of dosing, for instance, subdermal implantation of slow release formulations to give constant plasma concentrations, and intravenous injection to give very short pulses of high concentrations.

Caution is needed when dealing with the concept of the pseudo-steady state achieved after four days of dosing. There is no evidence for this in the Syngenta study. A relatively

constant level of DACT (the major metabolite) is achieved after 24 h, and is maintained throughout the dosing (see the various studies where individual metabolites are measured). The slow rise in total radio-label observed in the Thede (1987) study is difficult to explain. One possible explanation is that some binding is occurring over the four days, and that a pseudo steady state appears when all sites are saturated. However, it is difficult to see what such binding sites would be. If the pseudo steady state did involve binding, the AUC would be the sum of freely available (pharmacodynamically relevant) material and bound (unavailable to the site of action) material. If the rise in plasma concentration were to match a rise in tissue concentrations of free material, there is a need to identify the compartments into which the slow distribution process is taking place, and whether this contributes to a slow elimination process. It would be interesting to see the concentrations of DACT in the various body compartments over a period of days to find if there is any evidence of a slow accumulation in the tissues. In the Paul et al. (1993) data set, there is evidence that, following a single dose by oral gavage, there is a very slow elimination from liver and kidney, but an even slower elimination from the red blood cells (known to be due to covalent binding) and muscle. This is not matched by the rapid elimination from the plasma. It is difficult to gain information from the Simoneaux (1985) study since the first measurements were made on day 4 of the study, but the indications are that elimination from some tissues (including erythrocytes and muscle) was slow following the cessation of dosing (7 days of daily dosing).

The whole topic has become complicated because the time to pseudo steady state happens to be the same as the apparent critical exposure of 4 days over the rat estrous cycle. The two are independent of each other, and a 4 day exposure (at pseudo steady state or not) would have no effect in suppressing the LH surge if it were applied outside the critical time. If the rat were exposed for four days in the critical time, but before pseudo steady state was achieved, would a suppression of the LH surge be produced at the appropriate dose levels? The logic of assuming that a pseudo steady state is required would dictate that dosing would have to start four days before the onset of the critical exposure period in order to suppress the LH surge. As discussed above, there are now grounds for examining the idea that a constant level of atrazine is required over the whole 4 day critical period.

c) As pointed out in the Agency Issue Paper, we are still reviewing a PBPK model submitted by Syngenta. As we complete our review of the Syngenta model, please comment on key aspects that EPA should be considering concerning a PBPK model including model credibility (e.g., structure, parameter values, documentation), model reliability (e.g., how well does the model simulate the dose metrics relevant to the mode of action), and model applicability (e.g., does the model have essential features for intended application).

Panel Response:

As noted above, it is not an easy matter to evaluate or verify a complex model such as the PBPK model developed by Syngenta, and some of the key physiological aspects of the model need to be checked in order to establish creditability. These include the following:

- Physiological parameters (organ/tissue volumes and blood flows) should be realistic for each species and generally used in PBPK models.

- Major tissues and organs should be represented in the model, and arranged correctly with regard to anatomy and physiology.
- Tissue/blood (plasma) partition coefficients should be measured (not fitted).
- Metabolism and excretion parameters (clearances or rate constants) should be measured or estimated independently rather than obtained by fitting of the model.
- Partition coefficients, metabolism and excretion parameters should be independently assessed in each species for which the model is used; i.e., not measured in one species and used in another without independent confirmation that it is accurate for the second species.
- Tissues/organs that are sites of toxicological action should be present in the model.
- Toxicologically important species (active and contribute significantly to over-all toxic effect) should be included in the model.

Model reliability must also be established, and key aspects to be considered include the following:

- Using the same model structure and parameter values the model should describe the target organ/tissue and plasma concentration-time profiles for:
  - a broad range of doses that encompass those expected from environmental exposure
  - relevant modes of exposure; e.g., diet, drinking water, inspired air, skin exposure
  - acute and chronic exposures.
- The exposure metric(s) produced by the model should be consistent in a dose-response relationship with the apical endpoints.

Checks need to be made on the applicability of the model. A model should produce useful exposure metrics; e.g., AUC for free and total concentrations of toxicologically active species in plasma and at sites of the chemical initiating event in target organs/tissues. It is noteworthy that AUC for total concentration in target organs may not be adequate due to non-specific binding or accumulation in the target organ/tissue. Ideally, the model will provide an exposure metric based upon the free (unbound) concentration that is involved in the chemical initiating event at the site of the event.

It is important that the model can be used across the whole human population (including the wide range of adult body weights) and different life stages (including pregnant women and fetuses), and not just for an average adult human. Further, if the model is to be widely used, it should be available on transparent modeling platforms that are widely available to the modeling community and others who have an interest in applying the model.

Key aspects that EPA should be considering in their review of the PBPK model submitted by Syngenta, are addressed in the recently released World Health Organization guidance on the characterization and application of physiologically based pharmacokinetic (PBPK) models for application in risk assessment (WHO, 2010). Development of this guidance drew broadly on expertise internationally in both PBPK modeling and risk assessment and involved protracted input from a drafting group and in a series of related workshops. The working group developed a comprehensive list of questions for consideration relevant to evaluation of the biological basis, model simulations, reliability and applicability of specific

PBPK models for application in risk assessment. It is noted also that a subset of these questions is equally applicable to other types of pharmacokinetic models and it would be helpful to consider them, also, in relation to the one compartment modeling approach currently proposed by the Agency.

The WHO guidance document also makes recommendations concerning a process for consideration of PBPK models in regulatory risk assessment. This includes early and iterative involvement of regulatory risk assessors in model development, access to both internal and independent expertise, documentation by model developers in standard format for risk assessment applications and independent review (see the box below).

List of considerations for selecting PBPK models for use in risk assessment

Biological basis

- Are the major sites/processes of absorption, storage, transformation and clearance included in the model?
- Are the mathematical equations of ADME based on a sound theoretical or biological basis?
- Are the input parameters related to the characteristics of the host, chemical or environment?
- Is the sum total of the tissue blood flow rates equal to the cardiac output?
- Is the ventilation:perfusion ratio specified in the model within physiological limits?
- Are the volumes of compartments (individual and total) within known physiological limits?
- Is the approach used to establish partition coefficients within the domain of valid application?
- Is the method used for estimating biochemical parameters adequate?
- Is the allometric scaling of parameters, if applicable, done appropriately?
- Is the integration algorithm proven for solving differential equations in similar models?
- Has the computer model code been verified for syntax errors and the accuracy of units (i.e., dimensional consistency)?

Model simulation of data

- Has the model been evaluated for its ability to predict kinetics under various conditions, consistent with its intended application?
- Does the model consistently reproduce the general trend of the data (i.e., peaks, bumps and valleys, saturation of metabolism) or only portions of one or more data sets?
- Are the model predictions within an acceptable level of correspondence with the experimental data (e.g., within a factor of 2)?

Reliability (model testing, uncertainty and sensitivity)

- Is the model capable of providing predictions of the concentration time course of

the candidate dose metrics in the target organ or a suitable surrogate compartment (e.g., blood)?

- Has the uncertainty in model predictions of dose metric been assessed for the relevant exposure conditions?
- What is the reliability of the data used for calibrating and/or evaluating the PBPK model?
- Is the sensitivity of the dose metric to change in numerical values of input parameters characterized for relevant exposures?

Applicability

- Has the model been developed and evaluated in the species and life stage of relevance to the risk assessment?
- Do the exposure routes in the model correspond to those of anticipated human exposures as well as those of the critical studies chosen for the assessment?
- Has the model been tested for the exposure doses and durations of relevance to the intended extrapolations?
- Does the model contain point estimates (or distributions) of parameters, consistent with the purpose of application?

The PBPK model developed by the registrant has the potential to be an important tool that could provide the reliable, scientifically based extrapolation between species required by the Agency. This is feasible because the key physiological processes in the rat-based PBPK model are well understood and calibrated in humans. The PBPK model has some weaknesses that should be carefully evaluated (e.g., the use of *in vitro* metabolic studies to parameterize *in vivo* metabolism).

Although the concentration-time curves are well defined by frequent measurements in time, the predictions of tissue concentrations will depend heavily on the selected tissue-plasma partition coefficients. These values are critical and can introduce bias. It is reassuring that Syngenta intends to verify these *in vivo*. This may help to identify any binding within tissues, and check for the presence of a slow distribution compartment as discussed in responses to earlier sections in this charge question.

One current limitation is the inability, because of limits of quantification, to validate the model predictions for lower doses. It would be necessary to use a more sensitive method (e.g., radio-label) to assess performance over longer times and at lower doses. However, information on individual metabolites would then be lost. This should be considered by the Agency since risk assessments are generally based on low doses.

d) Please comment on the extent to which the one-compartment linear model of total plasma radioactivity derived from ^{14}C labeled atrazine may account for interspecies differences in pharmacokinetics.

Panel Response:

The one-compartment model adequately characterizes the pharmacokinetic behavior of radioactivity after administration of ^{14}C-atrazine to rat, monkey and humans. The available data indicate that the one-compartment model parameters are independent of the size of the dose and dose rate in each species over a broad range that encompasses the dose rate expected from some drinking water exposures. The parameter values generally differ among species within the bounds expected based upon allometric scaling principles, perhaps with the exception of monkey which shows ke values that are larger, not smaller as expected, than in rat. Values of volume of distribution per kg body weight (W) are generally body-weight independent and this appears to be the case for total plasma radioactivity. Systemic clearance values (CL, L/h) per kg W are generally expected to vary in proportion to $W^{-1/4}$, which makes the human CL value per kg about one-fourth the rat CL value per kg. As half-life is ln 2 * V / CL, the half-life in humans is expected to be about four times that in rat and this is approximately the case for plasma radioactivity. One compartment pharmacokinetic parameter values, therefore, appear to follow allometric scaling relationships generally observed for the pharmacokinetic parameter values of the majority of xenobiotics.

There is some uncertainty associated with the total radiolabel approach because of differences in metabolic capabilities between species. There are also interspecies differences in the sites and degree of binding (e.g., significant binding of DACT to erythrocytes in rat but not in other test species). These may be important since they play a part in determining the concentration of free toxicant in the plasma, and it is that which determines exposure of site of action (and of course, target tissues). Extrapolation can be effected using empirical allometric factors but it is not possible to carry out extrapolation on the basis of the scientific physiological and metabolic information. It is difficult to compare the exposure time needed for effects in rat with that in humans. Generally the information available from rat studies is based on the same dose being given on each day. This is not usual in humans where intake is sporadic in drinking water. This will become more important when extended exposure times (greater than 4 days) are involved. As discussed above, it was not clear why the rat did not scale to the monkey.

An advantage of the PBPK model is that it may lend itself to stochastic modeling studies of population pharmacokinetics and estimates of inter-individual variability (approach taken in Beaudouin et al., 2010, Bois et al., 2010). This sort of approach is not possible with studies based on total radiolabel.

Charge Question 13 - Any risk assessment of atrazine ideally should account for the temporal relationship between exposure and toxicological endpoint (i.e., LH attenuation). The Agency is relying on rat studies involving constant dosing frequency and dose level to elucidate this relationship. The time-to-effect in rats appears to be based on the build-up of plasma triazines to pseudo steady state (i.e., plasma triazine levels stay within a constant range). The Agency is using this information to estimate the allometrically scaled equivalent human time-to-pseudo steady state plasma triazine levels. However, the Agency is cognizant that human exposure is different from controlled rat studies in that both the dose level and the frequency of exposure through drinking water are variable. Thus, the Agency has identified several possible durations of human exposure, which collectively bracket the exposure intervals of concern for LH attenuation in humans:

o 28 days: a duration predicted by allometrically scaling the rat plasma elimination kinetics. In both rats and humans, the time estimated to reach pseudo steady state plasma triazines coincides with the respective length of their ovarian cycles (i.e., 4 and 28 days, respectively).

- 14 days: Proposed to serve as a midpoint between the other two durations for better characterization of water monitoring frequencies.

- 4 days: the duration of exposure needed to reach pseudo steady state in rats and corresponding to the follicular phase in the human menstrual cycle.

a) Please comment on the rationale used by the Agency for selecting these exposure duration options. Please discuss the rationale for other alternative durations of concern, if any.

Panel Response:

Time to reach steady state and time to effect are not necessarily closely related. It could be coincidence that the 4-day exposure in rat is both the time to pseudo steady-state of radioactivity and the duration of exposure that suppresses the LH surge at a low dose. That it is coincidence is supported by the observation in rat that the chloro-s-triazine (atrazine + deethyl-atrazine (DEA) + deisopropyl-atrazine (DIA) + diaminochlorotriazine (DACT)) exposure produced by the pseudo steady-state level is not much higher than that produced by the first dose. The Coder (2011b) study of four daily doses of atrazine by oral gavage followed by a four-day washout period, with plasma concentrations measured intensively during both the treatment and washout periods, showed very similar Cmax, Cmin and AUC for Treatment Days 2-4, for atrazine and its toxicologically active metabolites DEA, DIA and DACT. Treatment Day 1 exposures were only slightly smaller than those observed for Days 2-4. In other words, there was little accumulation of the chloro-s-triazines with a daily multiple dosing regimen, which is consistent with the relatively short half-lives of the chloro-s-triazines compared with the 24 hr dosing interval. For the 50 mg/kg/day oral gavage treatment (Coder 2011b), the longest effective half-life of the four triazines was that of DACT, which was about 7 hr. While the elimination of a very small fraction of DACT, starting 36 hr after the fourth dose, showed a half-life of about 17 hr, this half- life controls an insignificant fraction of the systemic exposure to

DACT. The 7 hr half-life and 24 hr dosing interval indicate that accumulation would be negligible (about 10% using the calculated accumulation ratio of $(1 - \exp(-KE*24))^{-1}$, where KE = 0.693 / 7 hr; i.e., at pseudo steady state, exposure would be 10% greater than after the first dose).

Since accumulation of the chloro-s-triazines is negligible when atrazine is dosed daily by oral gavage, the time to effect is apparently not controlled by the time required for the systemic concentrations of the triazines to reach a minimum critical level associated with the onset of effect, as the triazine exposure after the first daily dose is similar to that after the fourth dose. It is, therefore, more probable that the time to onset of effect is controlled by the pharmacodynamics; i.e., that the kinetics of events down-stream from the molecular-initiating event principally determine the onset of effect. The kinetics of the down-stream AOP events for LH attenuation in human vs. rat are not well characterized and it is, therefore, not apparent what the appropriate duration of human exposure is in the context of setting the maximum level of exposure to prevent LH attenuation in humans. Without the relative rat vs. human effect kinetics, the conservative (science policy-based) approach would be to use the 4-day duration identified in the studies with rats.

For total radioactivity plasma concentration, the elimination half-life is longer and the expected accumulation profile has a considerably longer time to steady state. In this case, the daily exposure would increase day by day, with 3-4 days of exposure required to achieve 90% of the steady-state plasma concentration, and accumulation to a threshold concentration could define the time to onset of the LH surge suppression. As the long half-life of total radioactivity likely reflects the half-life of albumin adducts, which are not active in LH surge suppression, this explanation of the 4-day exposure being defined by the time to reach steady state is unlikely.

The Panel did not believe they had scientific evidence to identify a critical period or length of exposure for humans. If the four day period based on the critical period in the rat estrous cycle is selected as a conservative, science policy-based option, information is not available to identify an equivalent critical period in humans.

b) Please comment on which exposure duration in humans most closely corresponds to the exposure duration found to cause adverse effects in rats.

Panel Response:

An important factor to consider is the minimum duration of LH attenuation that must occur before adverse toxicological effects ensue. Is a brief, transient suppression of LH to be avoided, or suppression of longer duration? How large a suppression of the LH surge must be avoided? Currently, sufficient information is not available to answer these questions. Since critical information in this area is lacking, the conservative science policy-based approach is to avoid even a brief, transient suppression of LH, and the four-day exposure identified for rat would therefore seem appropriate.

c) Please comment on the approach used by the Agency (i.e., the one-compartment linear model) to relate atrazine levels from the water chemographs to predict corresponding human plasma triazine levels for the proposed durations of concern. In particular, please comment on the Agency's proposed approach to use water AUC estimates to calculate a time-weighted daily average of atrazine exposure for a given duration of concern. Please suggest alternative approaches as appropriate.

Panel Response:

The approach is theoretically sound, and may benefit from the use of the Syngenta data since these are based on a much larger number of time points than any of the other studies. Those data will be more reliably (less dependent on empirical transformations) extrapolated to humans. The approach does provide an objective way of moving from rodent to human. The PK approach and data will be reliable, but the toxicological assumptions need to be examined in light of the recent data that have emerged.

The integral of the water chemograph divided by the time span of the chemograph provides an estimate of the time-weighted average concentration of triazines in the water during the time span. Multiplication of the average triazine concentration by the daily water ingestion rate quantifies the daily triazine dose. The daily triazine dose divided by the human triazine clearance (CL = ke*Vdss* = 0.004 h^{-1} * 24 hr/day * 6.55 L/kg * 60 kg bw = 38 L/d) provides an estimate of the steady-state average plasma concentration of total atrazine and metabolites (measured using measurements of total radioactivity), and when that is multiplied by 24 hr, the AUC for the human over a 24 hr period at steady state is obtained.

The use of water AUC to calculate a time-weighted daily average atrazine exposure is theoretically sound. This provides the time weighted average (TWA) concentration to which humans are exposed (providing that it is finished water), and hence, a reasonable metric of exposure. It will, of course, flatten peak events, but depending on the nature of the water storage, treatment and holding capacity for treated water, this may be a reasonable reflection of reality when considering water at the tap. It seems unlikely that humans would ever experience the sorts of internal exposures necessary in rats to produce suppression of the LH surge. It is more likely to be a prolonged exposure to low concentrations, and so the TWA approach seems the most reasonable. An alternative to the water AUC is simply to average the measured water concentrations, but this would be inferior to using the AUC. A way to "fill in the gaps", when they occur, using rainfall and/or stream flow measurements coupled with historical water concentration values would be helpful.

Charge Question 14 - For the case study, the Agency used the 95th and 5th percentile of conditional simulations of daily concentrations from variogram models based on 7-day sampling intervals to predict human plasma AUC for triazines. The estimations of daily concentrations from weekly sampling were used to calculate the rolling average concentrations for the potential critical window of exposure from 4 to 28 days. The simulation models underestimated a single day peak concentration but appear to provide reasonable approximations of rolling average concentrations with durations as short as 4 days as well as the length of time of the potential exceedance.

a) Please comment on the use of a 95th percentile of the conditional simulations for providing an upper bound on rolling average concentrations in the case study.

Panel Response:

The consensus of the Panel was that all the statistical issues related to exposure presented in response to this Charge Question as well as Charge Questions 1-4 are manageable and much easier to deal with than the biological or mechanistic issues that have been discussed. In particular, properly computing and interpreting the 95th percentile of the conditional simulations requires a clearer understanding and delineation of sources of error.

Identified sources of error noted in the reports include: sample size (to which sampling error links), spatial and temporal proximity of samples (which alludes to coverage, and hence quality, of samples; note: spatial proximity does not appear to be used), and the nature of a given phenomenon (i.e., its inherent variability). Model misspecification should be added to this list, as should measurement error. This latter source of error can be linked to the substitution of kriged or deterministic model-generated imputations into a daily time series, as well as the handling of below-detection-limit values. Although it furnishes a tool to ascertain uncertainty and risk, conditional simulation (which utilizes Monte Carlo techniques) does not embrace all of these sources of error. One weakness is that conditional simulation is sensitive to the data upon which conditioning is made: a simulation replicates its conditioning values, on average. Frequently the normal or log-normal distribution is the probability model of choice that is attached to the conditioning values; in other words, the conditioning values may be the means of a collection of normal distributions, one for each day in a time series. For the specimen atrazine data, a log-normal distributional assumption fails to adequately track the serial correlation in the data, and furnishes a poorer statistical description for the monitoring data than selected alternative statistical distributions.

When considering a 90% confidence interval (i.e., 5^{th} and 95^{th} percentiles), a balance should be maintained between claiming an excessive atrazine concentration when one does not exist, and failing to detect an excessive atrazine concentration when one does exist. The latter is the riskier of these two situations. For the purpose of atrazine impact analysis, this confidence interval focuses attention on the 95^{th} percentile from a conditional simulation. Because rolling average concentrations are means, by definition they result from smoothing data so that peaks disappear (consequently, actual peak concentrations are underestimated). These averages are easier to predict specifically because they are means. Figures 27 and 29 in the EPA Issue Paper illustrate that 14- and 28-day rolling averages may be of little value for decision-making and

monitoring purposes, even though they have relatively tight confidence intervals; although 4-day rolling averages are better, the crucial peak missed by them is a substantial peak. This point may be of less importance if the rolling averages represent duration of exposure.

The situation may well improve with a change in the variogram model as well as a change in the probability model employed for the stochastic simulation. A variogram model that better captures autocorrelation effects will better differentiate between the conditional and stochastic components. A probability model that allows more variability has the potential to better capture peaks.

A standard conditional simulation fails to capture all sources of variation. It assumes that the conditioning (e.g., kriged values) are fixed and true, and then simulates sampling error about these values. It fails to incorporate parameter estimation error. The log-transformed atrazine example time series data can be used to compute parameter uncertainty (see Table 5)

Table 5. Parameter estimates and related standard errors for selected variogram forms.

parameter	spherical		Gaussian		Bessel	
	estimate	s.e.	estimate	s.e.	estimate	s.e.
Nugget	0.00	*	0.02	0.01	0.01	0.01
Scale	1.72	0.10	1.66	0.08	1.73	0.12
Range	21.68	1.79	8.01 (eff r = 13.9)	0.49	9.27 (eff r = 37.1)	1.54
Nu					2.80	2.56
WSSE	28.1		13.2		12.9	
AIC	4.0		-18.6		-17.3	

The standard conditional simulation also fails to incorporate measurement error. The substitution of selected quantities for below-detection limit values should have minimal impact upon results. The variability in an assay to detect and quantify atrazine in water samples can be more substantial. RaPID Assay Kit analytical precision standards are $\pm$ 30%, although laboratory analysis by gas chromatography/mass spectrometry (GCMS) is substantially more precise. The cumulative effect of these errors (i.e., at least some of them may compound), which propagate through an analysis, may well invalidate the claim of a 95th percentile. Without tracing the cumulative effects of these different sources of error, perhaps a more representative approach would be to use the 97.5th percentile (i.e., a 95% confidence interval).

Finally, the method illustrated for computing the 90% upper and lower confidence bounds for AUCwater are inconsistent with standard statistical definitions of these quantities and how the conditional simulations are run. The AUCwater value is a time series that is computed from an input water concentration times series by successively integrating numerically (using the trapezoid method) over the duration of concern (say four days). This derived time series is then used as input to compute the AUCplasma. The time step for the underlying concentration time series is daily, and the resulting time step for the AUCwater time series is daily. The goal of this exercise is the estimation of human average daily concentration of atrazine over a yet unspecified period of concern. For the 4-day duration of concern, concentrations from 4 sequential days (five actual time points) are input into the trapezoid method to produce an average AUCwater

value. For a 14 day duration of concern, concentrations from 14 sequential days (15 points) are input into the trapezoid method to produce an average AUCwater (see Figure 21 in the EPA Issue Paper). The first period of concern uses concentrations from days 1 to 5, the second period uses days 2 to 6, etc. Thus, if we have a total of T days of data, for the 4-day duration of concern, the resulting time series will consist of T-4 time points assigned to days t=5 to t=T.

Consider first the Daily Sampling concentration time series. Figure 29a in the EPA Issue Paper shows an actual Daily Time Series curve, a Simulated 95th percentile curve and a Simulated 5th percentile curve. Passing the Actual Daily Sampling time series through the numerical integration function produces an "actual" AUCwater times series that then gets passed through a second one-to-one function to produce the ACSplasma, or "actual" plasma concentration time series.

Exactly how the 95th percentile curve from Figure 29a is translated into an upper AUCwater time series is not clear from Chapter 7 in the EPA Issue Paper. The remaining discussion is based on the assumption that the proposed method is to take the time series represented by the simulated 95th percentile curve and pass it through the numerical integration function to produce the "95th percentile ACSwater time series." Something similar was done with the 5th percentile concentration curve to produce the "5th percentile AUCwater time series."

This approach is not the statistically correct way of computing the 95th or 5th percentile distributions for the AUCwater time series. The proper method is to start with the original 1,000 individual simulations of the concentration series and compute an AUCwater time series for each of the simulated concentration time series. From these 1,000 AUCwater time series, one can now estimate the 95th and 5th percentiles of the AUCwater values for each time point by sorting the simulated values at each point and using as estimates the 5th and 95th percentiles of this distribution.

This process is illustrated in Figures 1-5 below. [Note: Subsequent to the meeting, the Panel member who produced these images realized that what were presented at the meeting as upper and lower curves were computed as single day max and min values instead of 5th and 95th percentile values. The original figures have been modified to correct for this, and one additional figure added that was part of the discussion but not originally presented. This does not change the original point the Panel member was attempting to make, namely that the method suggested in Chapter 7 of the EPA Issue Paper would represent a very conservative bound on concentrations.]

In Figure 2, the shaded area represents 1,000 simulated concentration time series. Computing the 5th and 95th percentiles across the 1,000 simulated values at each day produces the upper and lower curves inside the scatter produced in Figure 3. Running these three times series through the AUCwater equation, assuming a 4-day period of interest, produces the curves given in Figure 4. It is important to note how the AUC averages the original concentration and tends to dampen the peaks and valleys of the original time series. In Figure 5, the AUC times series computed for the individual 1,000 simulated concentration time series have been computed and plotted as individual lines. Note that these 1,000 simulated AUC curves almost fit entirely within the computed 5th and 95th percentile curves of Figure 4 that have been over

plotted on the 1,000 AUC curves. The conservativeness of using the 5th and 95th concentration curves in the AUC calculation directly is shown by the fact that they more closely represent minimum and maximum simulated AUC values for each day. Figure 6 illustrates the proper 5th and 95th percentile AUC curves. As would be expected, they fall inside the scatter.

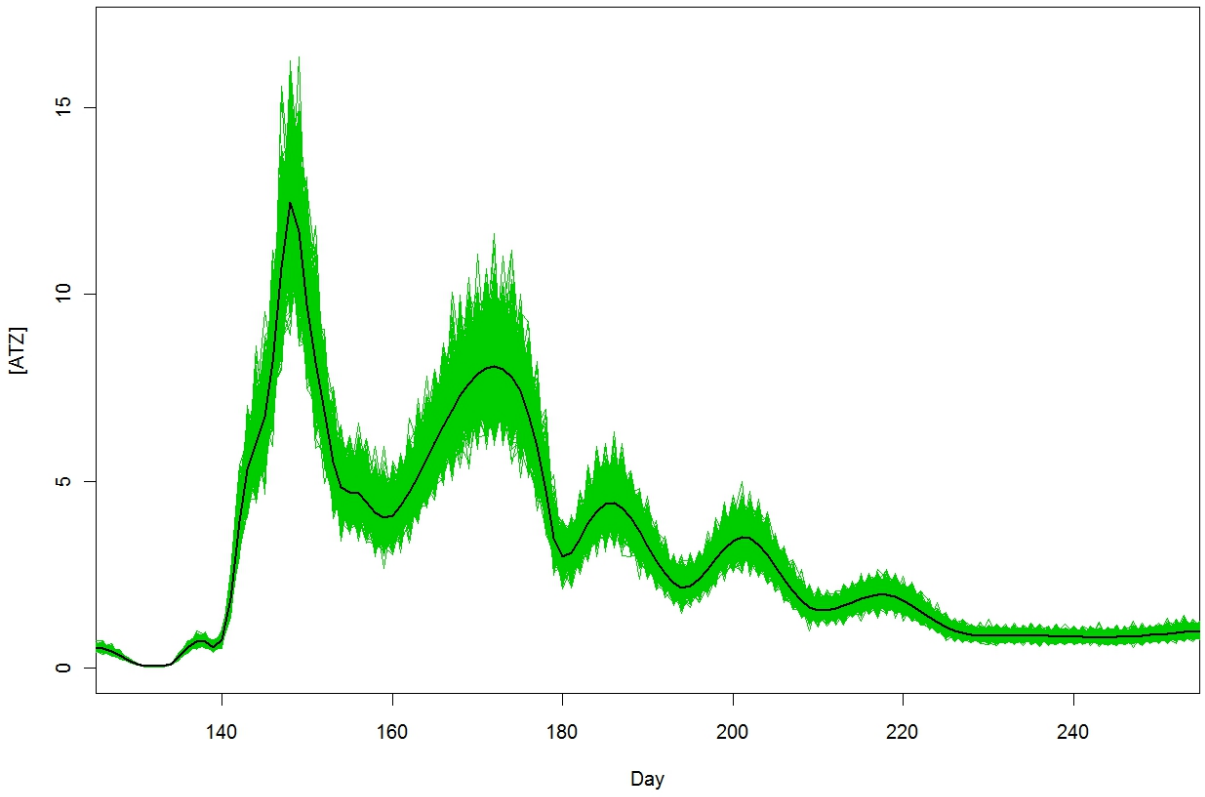


Figure 2: 1,000 simulated atrazine concentration time series (green lines) with the actual average time series presented in the middle of the scatter (black line). The average time series was modeled on a smooth in-fill of the 1995 Maumee River daily data. Note that the individual simulated time series vary around the mean line.

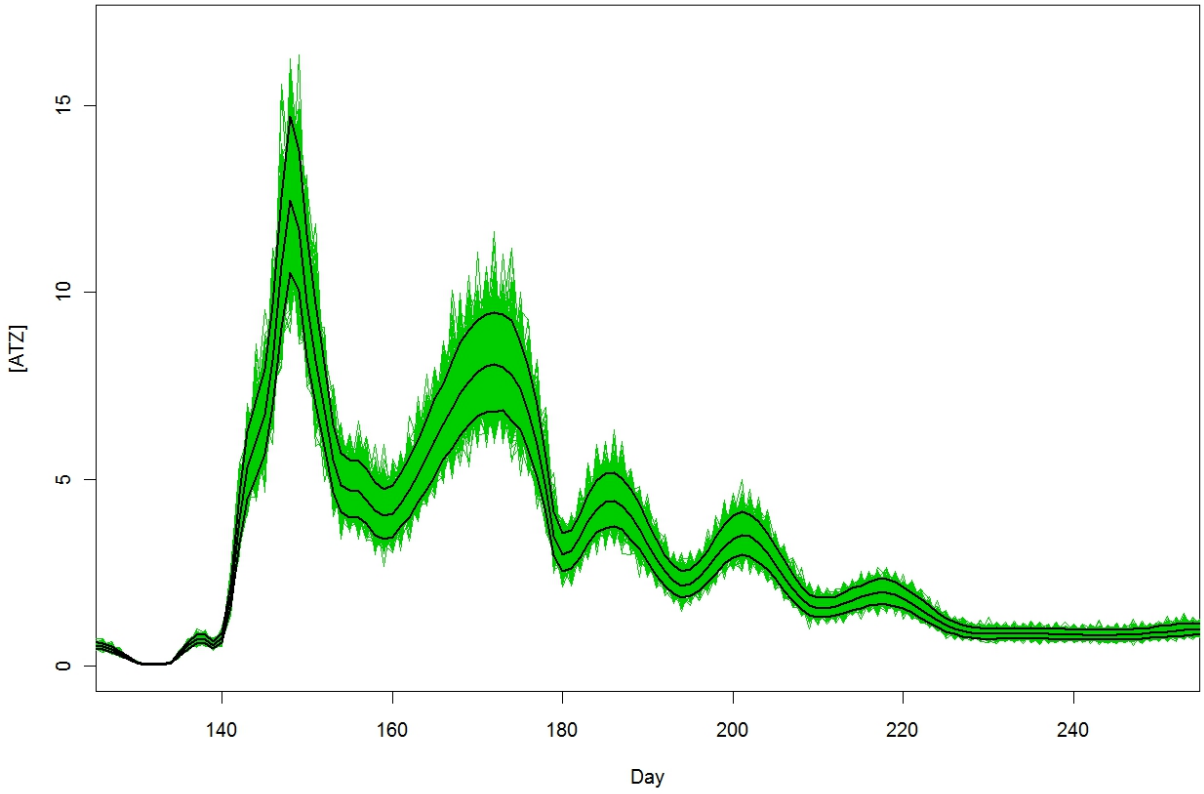


Figure 3: Same data for Figure 2, only now the time series that connects the individual 95th percentile values of the individual time series and the line that connects the 5th values of the individual time series are included.

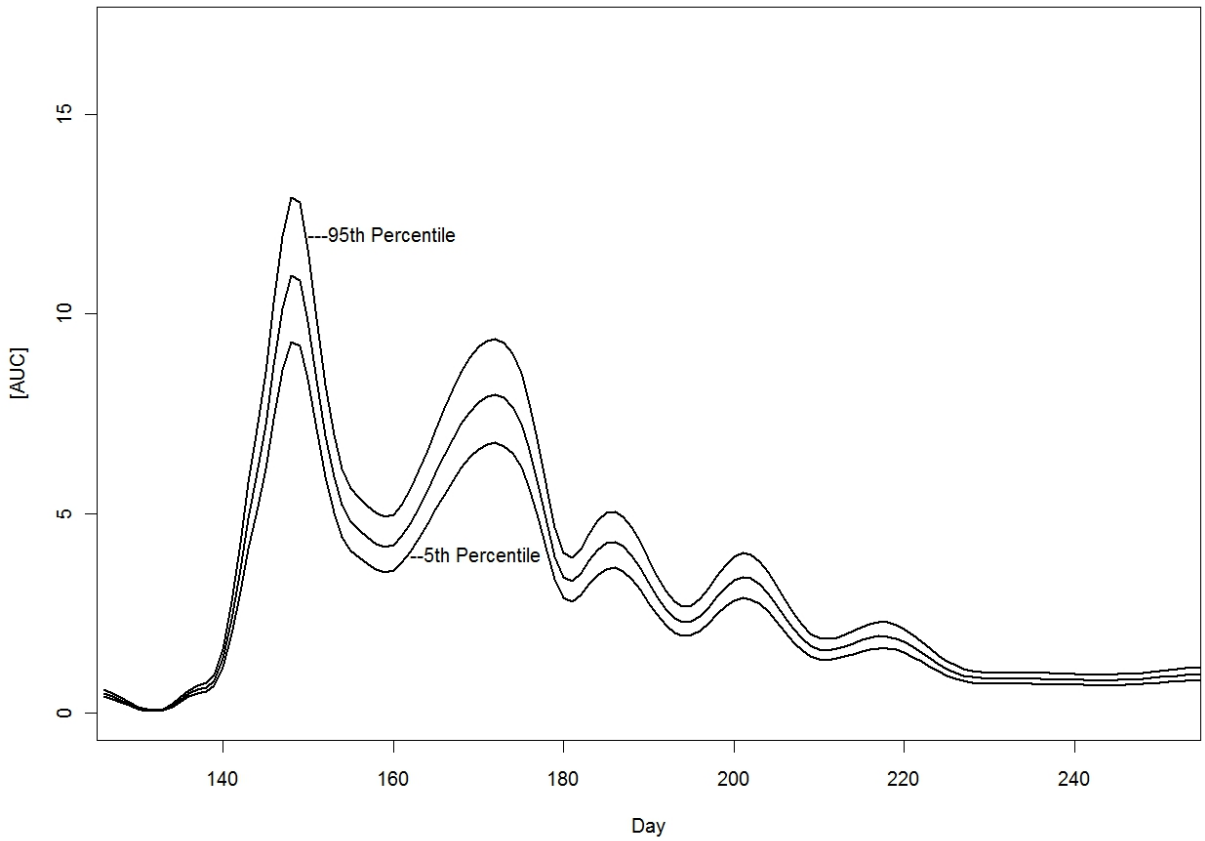


Figure 4: Estimated AUCwater values computed by running the values for the 5th, mean and 95th simulated time series lines from Figure 3 through the AUCwater equation, assuming a 4-day period of interest. The Trapezoid method uses 5 sequential concentration values to produce one AUC value which, in this case, is plotted at the mid-point of the interval.

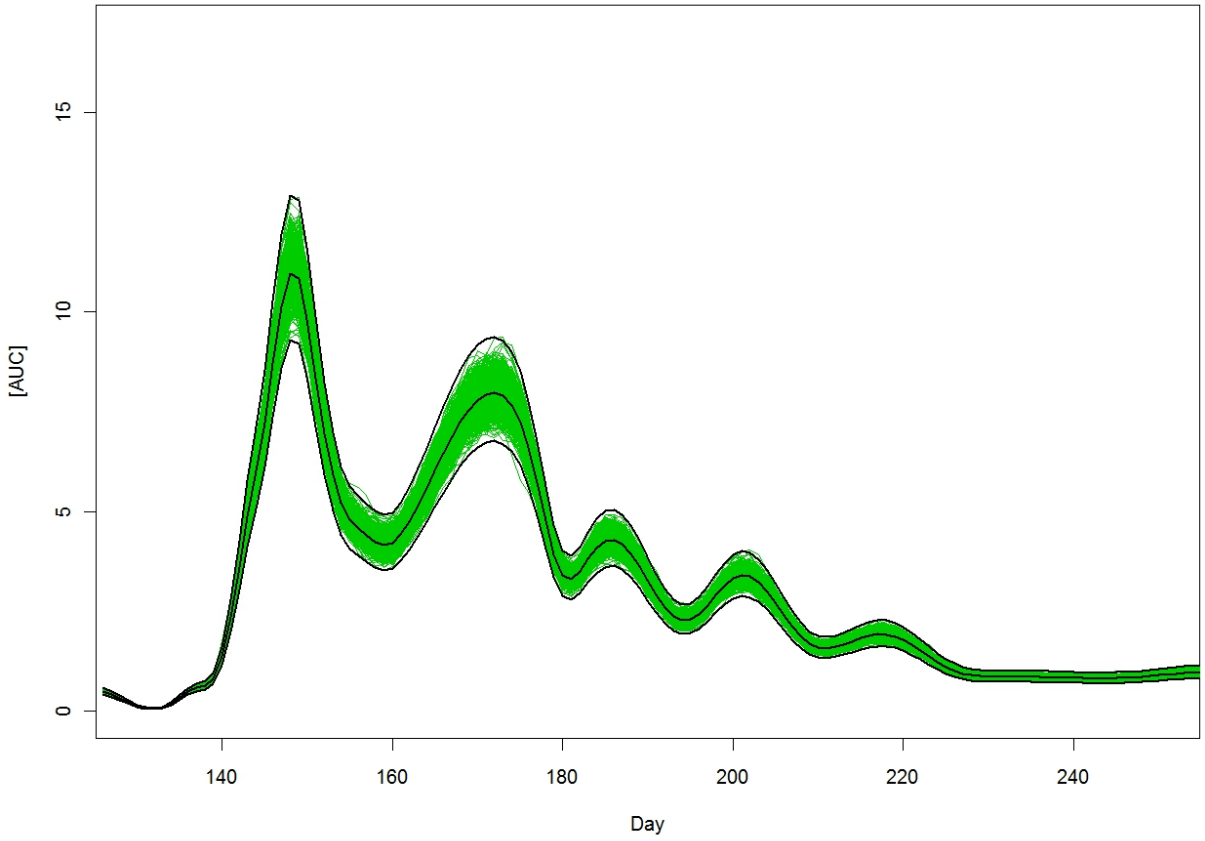


Figure 5: The estimated AUCwater time series lines computed for each of the 1,000 simulated times series are overplotted along with the curves from Figure 4. Note how the 5th and 95th time series curves bound almost all the individual simulated AUCwater time series. The 5th and 95th AUCwater time series are overestimates, being better predictors of the worst and best case AUCwater time series than providing a 90th confidence interval.

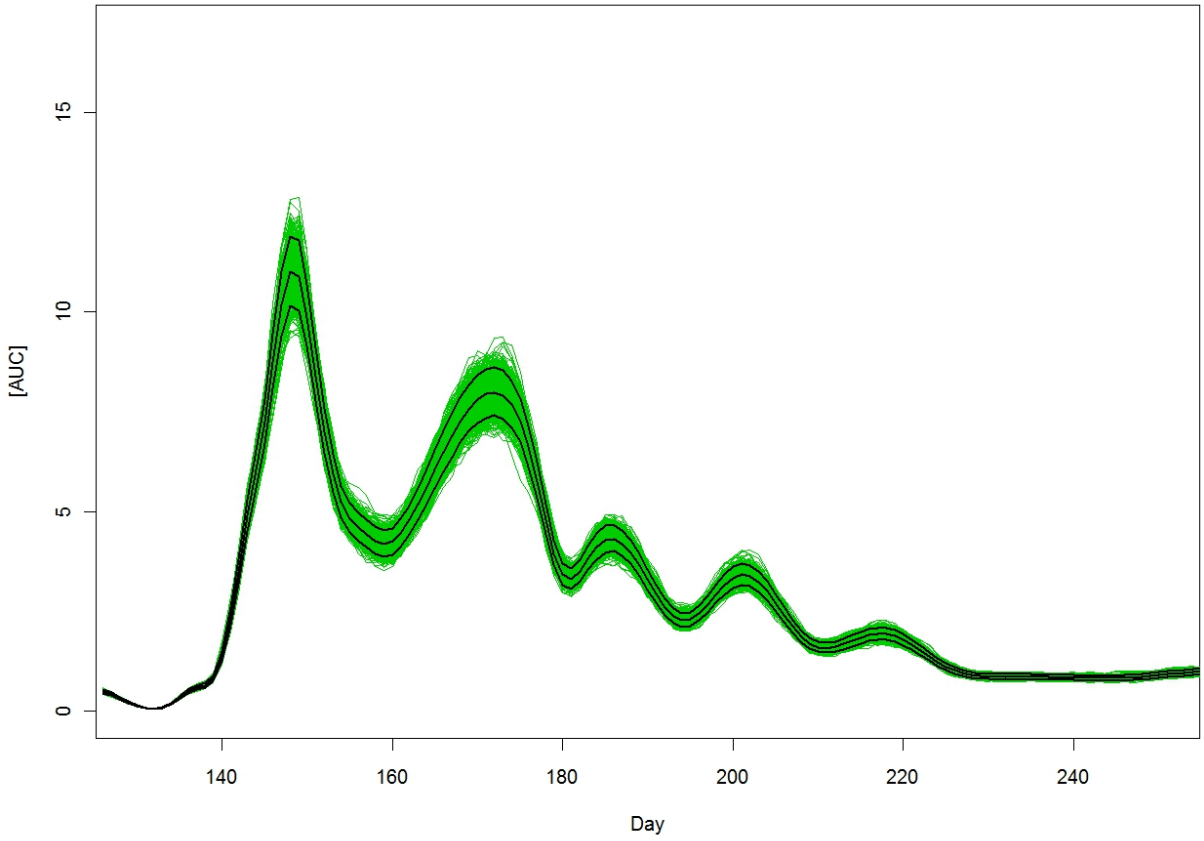


Figure 6: Over the estimated AUCwater time series lines for each of the 1,000 simulated times series have been plotted the 5th and 95th time series computed from the 1,000 values for each day. These would be the proper series to use if a 90th percent confidence band for AUCwater were to be used in the risk assessment.

During the Panel's deliberations, there was occasional reference to the effect of averaging on the daily concentration time series. Figure 7 illustrates this impact by displaying the effect of different averaging times on the time series. In order that the averaging does not shift the series dramatically, the mid-point of each averaging period is used for the X-axis point in plotting. Note that at some point, in this case at 28-day averaging, not only is peak concentration reduced, but peaks begin to be combined and the expected day of maximum peak is shifted.

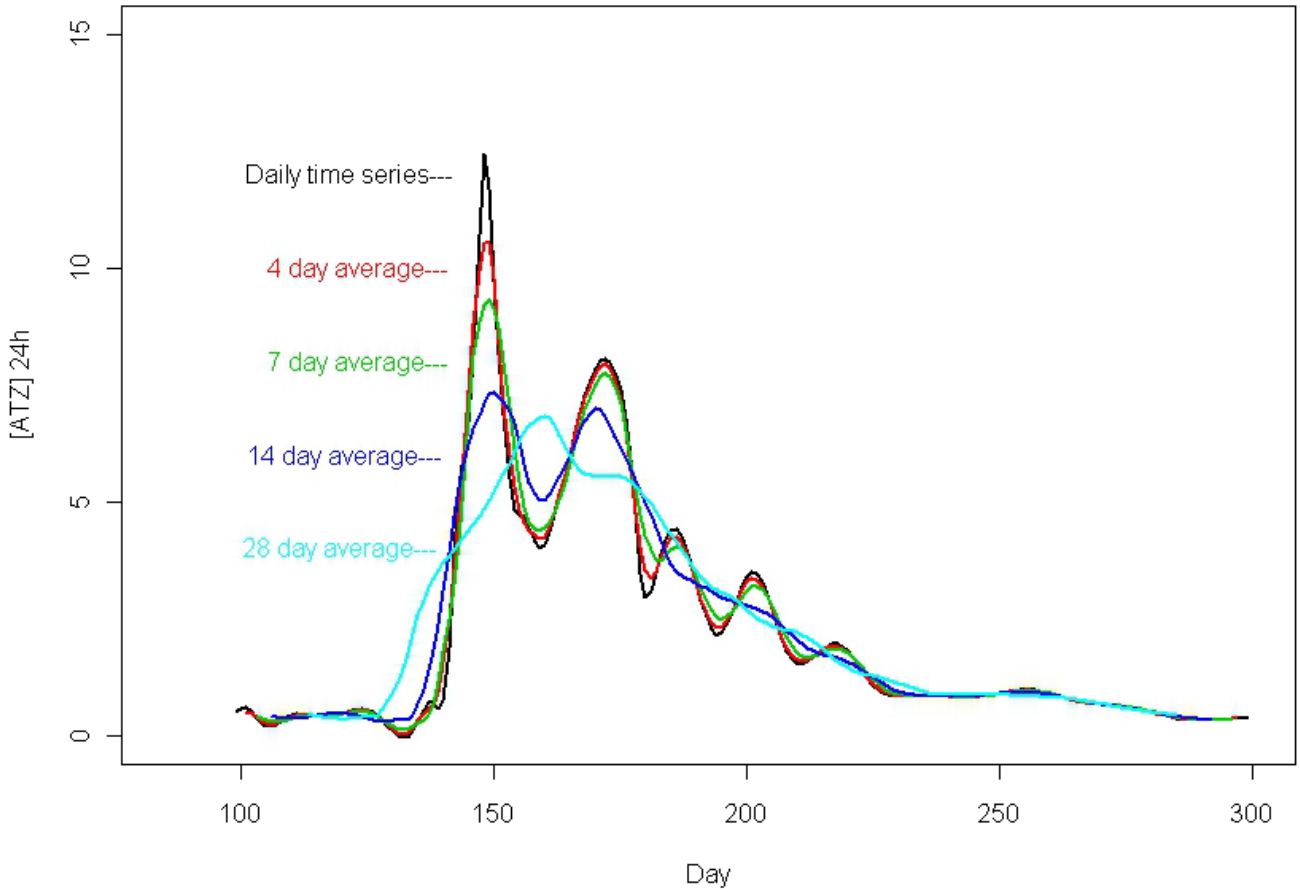


Figure 7: Using the daily time series modeled on a smooth in-fill of the 1995 Maumee River daily data, 4-, 7-, 14- and 28-day average time series are displayed. Note that as the averaging period increases, the details of the "concentration events" are lost.

In one aspect, computation of the AUCwater term using the Trapezoid rule produces exactly the same effect as time averaging of concentrations. The Trapezoid rule is actually a type of average. In Figure 8, AUC values for different durations are presented. Note that the concentration presented on the Y-axis is a 24-hour total, not the per-hour value expected for subsequent risk assessment. In addition, the AUC value is plotted at the mid-point of the averaging period to allow peaks to line up with the original single day peaks.

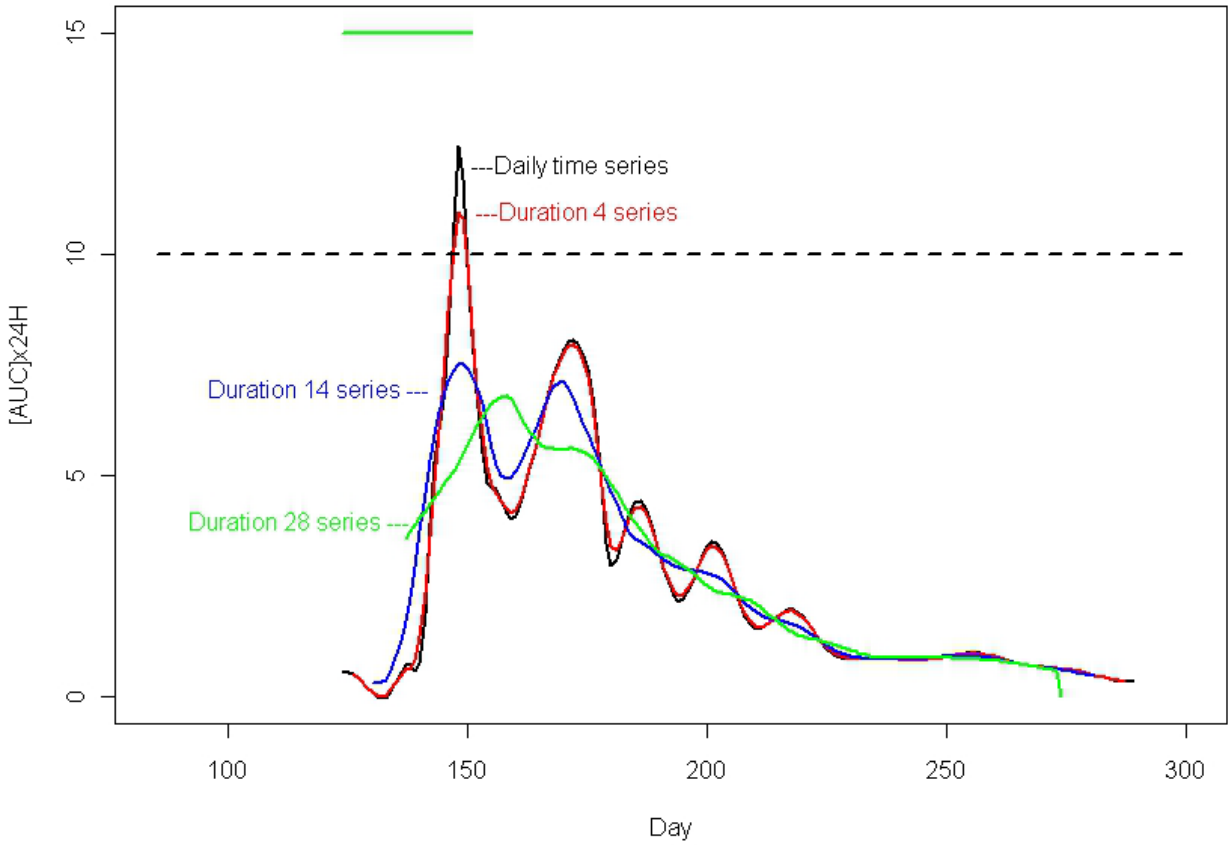


Figure 8: The transformation from daily AUCwater time series to series of increasing duration produces similar curves to period averaging. This is because with equal time spacings, the trapezoid numerical interpolation is actually very similar to the period average.

Additional Editorial Revisions Suggested for EPA Issue Paper and Appendices

- Some references are noted in the text but do not appear in the bibliography Attachment 2 (Cancer Epidemiology), specifically: Alavanja 2005, Blair 2011, Beane Freeman 2011, Flower 2004, Koutros 2010, Waggoner 2011.
- At the top of page 35, the reader is referred to Appendix C for the methodology for the literature review, but it seems that this description actually appears in Appendix B.
- Page 36 makes reference to Appendix F, which could not be located in the docket.
- In section 3.2.3.1., first paragraph on page 41: "quintiles" and "quartiles" appear to be used interchangeably, though they are clearly not interchangeable, but it appears that what is intended is "quartiles", so "quintiles" should be changed to "quartiles" in all occurrences in this paragraph.
- In the first line of the second complete paragraph on page 49, "evaluation" should not be plural.
- In the fifth line of the third paragraph of section 3.2.3.4, thyroid cancer, reference is made to "…the small number of atrazine exposed cases in some categories, e.g., n>5…"; presumably the ">" sign should be "<". Also, the last sentence of this section on thyroid cancer on page 53 does not appear to be a complete sentence.

- The first paragraph on page 54 needs a few edits: a) in the fourth line "cancers" should be inserted after "…cases of lymphohematopoietic…"; b) a bit more detail should be provided in this paragraph about the methods used to identify controls "from the general population"; and c) in the penultimate sentence of this paragraph, "only" should be moved to be inserted before "Hoar et al.". Additionally, in the first line of the last paragraph on this page, "studies" should be inserted after "…NCI sponsored". Also, in the last line of the second paragraph on page 55, "are noted" should be deleted.
- In the sixth line on page 67, "studies" should be "study".

REFERENCES

Alavanja MC, Samanic C, Dosemeci M, Lubin J, Tarone R, Lynch CF, et al. 2003. Use of agricultural pesticides and prostate cancer risk in the Agricultural Health Study cohort. *Am J Epidemiol* 157:800-814.

Barry MJ, McNaughton-Collins M. 2007. Benign Prostate Disease and Prostatitis. In Goldman: *Cecil Medicine*, 23rd Ed., Chapter 130.

Beane Freeman LE, Rusiecki JA, Hoppin JA, Lubin JH, Koutros S, Andreotti G, Zahm SH, Hines CJ, Coble JB, Barone-Adesi F, Sloan J, Sandler DP, Blair A Alavanja MC. 2011. Atrazine and cancer incidence among pesticide applicators in the Agricultural Health Study (1994-2007). Environ Health Perspect 119(9):1253-9.

Beaudouin R, Micallef S, Brochot C. 2010. A stochastic whole-body physiologically based pharmacokinetic model to assess the impact of inter-individual variability on tissue dosimetry over the human lifespan. *Regul Toxicol Pharmacol* 57:103-16.

Begley LA, Kasina S, MacDonald J, Macoska JA. 2008. The inflammatory microenvironment of the aging prostate facilitates cellular proliferation and hypertrophy. *Cytokine* 43(2):194-9.

Blair A, Stewart P, Lubin JH, Forastiere F. 2007. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. Am J Ind Med. 50(3):199-207.

Blair A, Thomas K, Coble J, Sandler DP, Hines CJ, Lynch CF, Lubin JH. 2011. Impact of pesticide exposure misclassification on estimates of relative risks in the Agricultural Health Study. Occup Environ Med. doi: oem.2010.059469[pii] 10.1136/oem.2010.059469.

Bois FY, Jamei M, Clewell HJ. 2010. PBPK modeling of inter-individual variability in the pharmacokinetics of environmental chemicals. *Toxicology* 278:256–267.

Bowen JM, Dahl GE, Evans NP, Thrun LA, Wang Y, Brown MB, Karsch FJ. 1998. Importance of the Gonadotrophin-Releasing Hormone (GnRH) Surge for Induction of the Preovulatory Luteinizing Hormone Surge of the Ewe: Dose-Response Relationship and Excess of GnRH. *Endocrinology* 139(2):588-595.

Brown LM, Blair A, Gibson R, Everett GD, Cantor KP, Schuman LM, Burmeister LF, Van Lier SF, Dick FR. 1990. Pesticide exposures and other agricultural risk factors for leukemia among men in Iowa and Minnesota. *Cancer Res* 50(20):6585-6591.

Brown LM, Burmeister LF, Everett GD, and Blair A. 1993. Pesticide exposures and multiple myeloma in Iowa men. *Cancer Causes Control* 4(2):153-156.

Cantor KP, Blair A, Everett G, Gibson R, Burmeister LF, Brown LM, Schumann L, Dick FR. 1992. Pesticides and other agricultural risk factors for non-Hodgkin's lymphoma among men in Iowa and Minnesota. *Cancer Res* 52(9):2447-2455.

Carreon T, Butler MA, Ruder AM, Waters MA, Davis-King KE, Calvert GM, Schulte PA, Connally B, Ward EM, Sanderson WT, Heineman EF, Mandel JS, Morton RF, Reding DJ,

Rosenman KD, Talaska G. 2005. Gliomas and farm pesticide exposure in women: The Upper Midwest Health Study. *Environ Health Perspect* 113(5):546-551.

Clavel J, Hemon D, Mandereau L, Delemotte B, Severin F, and Flandrin G. 1996. Farming, pesticide use and hairy-cell leukemia. *Scand J Work Environ Health* 22(4):285-293.

Cooper RL, Laws SC, Das PC, Narotsky MG, Goldman JM, Lee Tyrey E, Stoker TE. 2007. Atrazine and reproductive function: mode and mechanism of action studies. *Birth Defects Res B Dev Reprod Toxicol* 80(2): 98-112.

Davis LK, Murr AS, Best DS, Fraites MJP, Zorrilla LM, Marotsky MG, Stoker TE, Goldman JM, Cooper RL. 2011. The effects of prenatal exposure to atrazine on pubertal and postnatal reporductive indices in female rat. *Reprod Toxicol* doi:10.1016/j.reprotox.2011.04.004.

De Marzo AM, Marchi VL, Epstein JI, Nelson WG. 1999. Proliferative inflammatory atrophy of the prostate: implications for prostatic carcinogenesis. *Am J Pathol* 155: 1985-92.

De Marzo AM, Platz EA, Sutcliffe S, Xull J, Grönberg H, Drake CG, Nakai Y, Isaacs WB, Nelson WG. 2007. Inflammation in prostate carcinogenesis. *Nature Rev Cancer* 7: 256-69.

De Roos A, Zahm S, Cantor K, Weisenburger D, Holmes F, Burmeister LF, Blair A. 2003. Integrative assessment of multiple pesticides as risk factors for non-Hodgkin's lymphoma among men. *Occup Environ Med* 60(9):E11.

Donna A, Betta PG, Robutte F, Crosignani P, Berrino F, Bellingeri D. 1984. Ovarian mesothelial tumors and herbicides: a case-control study. *Carcinogenesis*. 5(7):941-942.

Donna A, Crosignani, P, Robutti, F, Betta, P, Bocca, R, Mariani, N, Ferrario F, Fissi R, Berrino F. 1989. Triazine herbicides and ovarian epithelial neoplasms. *Scand J Work Environ Health* 15(1):47-53.

Engel LS, Hill, DA, Hoppin, JA, Lubin, JH, Lynch, CF, Pierce, J, Samanic C, Sandler DP, Blair A, Alavanja MC. 2005. Pesticide use and breast cancer risk among farmers' wives in the agricultural health study. *Am J Epidemiol* 161(2):121-135.

Enoch RR, Stanko JP, Greiner SN, Youngblood GL, Rayner JL and Fenton SE. 2007. Mammary gland development as a sensitive endpoint after acute prenatal exposure to an atrazine metabolite mixture in female Long-Evans rats. *Environ Health Perspect* 115(4)541-547.

Everett JW, Krey LC, Tyrey L. 1973. The Quantitative Relationship Between Electrochemical Preoptic Stimulation and LH Release in Proestrous Versus Late-Diestrous Rats. *Endocrinology* 93:947-953.

Flower KB, Hoppin JA, Lynch DF, Blair A, Knott C, Shore DL, Sandler DP. 2004. Cancer risk and parental pesticide application in children of Agricultural Health Study participants. *Environ Helath Perspect* 112(5):631-635.

Foradori CD, Hinds LR, Hanneman WH, Handa RH. 2009a. Effects of Atrazine and Its Withdrawal on Gonadotropin-Releasing Hormone Neuroendocrine Function in the Adult Female Wistar Rat. *Biol Reprod* 81(6):1099-1105.

Foradori CD, Hinds LR, Hanneman WH, Legare ME, Clay CM, Handa RJ. 2009b. Atrazine Inhibits Pulsatiel Luteinizing Hormone Release Without Altering Pituitary Sensitivity to a Gonadotropin-Releasing Hormone Receptor Agonist in Female Wistar Rats. *Biol Reprod* 81(1):40-45.

Goldman J, Davis L, Murr A, Cooper R. 2011. The influence of 1, 2, and 4 days of atrazine treatment on the LH surge of ovariectomized/estradiol-primed rats: Preliminary report to the Office of Chemical Safety and Pollution Prevention. Internal Report Submitted 28 April 2011. National Health and Environmental Effects Research Laboratory, EPA Office of Research and Development.

Herbison AE, Porteous R, Pape JR, Mora JM, Hurst PR. 2008. Gonadotropin-releasing hormone neuron requirements for puberty, ovulation, and fertility. *Endocrinology* 149(2):597-604.

Hessel PA, Kalmes R, Smith TJ, Lau E, Mink PJ, Mandel J. 2004. A nested case-control study of prostate cancer and atrazine exposure. *Occup Environ Med* 46(4):379-85.

Hoar SK, Blair A, Holmes FF, Boysen C, Robel RJ. 1985. Herbicides and colon cancer. *Lancet* 1(8440): 1277-1278.

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

Hopenhayn-Rich C, Stump M, Browning S. 2002. Regional assessment of atrazine exposure and incidence of breast and ovarian cancers in Kentucky. *Arch Environ Contam Toxicol* 42(1):127-136.

Laws SC, Ferrell JM, Stoker TE, Schmid J and Cooper RL. 2000. The effects of atrazine on female Wistar rats: an evaluation of the protocol for assessing pubertal development and thyroid function. *Toxicol Sci* 58(2):366-376.

Lee WJ, Sandler DP, Blair A, Samanic C, Cross AJ, Alavanja MC. 2007. Pesticide use and colorectal cancer risk in the Agricultural Health Study. *Int J Cancer* 121(2):339-346.

MacLennan PA, Delzell E, Sathiakumar N, Myers SL, Cheng H, Grizzle W, Chen, VS, Wu XC. 2002. Cancer incidence among triazine herbicide manufacturing workers *Occup Environ Med* 44(11):1048–1058.

MacLennan P, Delzell E, Sathiakumar N, Myers S. 2003. Mortality among triazine herbicide manufacturing workers. *J Toxicol Environ Health* A. 66(6):501-517.

McElroy J, Gangnon R, Newcomb P, Kanarek M, Anderson H, Brook J, Trentham-Dietz AMY, Remington PL. 2007. Risk of breast cancer for women living in rural areas from adult exposure to atrazine from well water in Wisconsin. *J Expo Sci Environ Epidemiol* 17(2):207-214.

Mills PK. 1998. Correlation analysis of pesticide use data and cancer incidence rates in California counties. *Arch Environ Health* 53:410-413.

Mills PK, Yang R. 2003. Prostate cancer risk in California farm workers. *Occup Environ Med* 45:249-258.

Mills P, Yang R. 2006. Regression analysis of pesticide use and breast cancer incidence in California Latinas. *J Environ Health* 68(6):15-22; quiz 43-14.

Muir K, Rattanamongkolgul S, Smallman-Raynor M, Thomas M, Downer S, Jenkinson C. 2004. Breast cancer incidence and its possible spatial association with pesticide application in two counties of England. *Public Health*. 118(7):513-520.

Muntzing J, Sufrin G, Murphy GP. 1979. Prostatitis in the rat. *Scand J Urol Nephrol* 13:17-22.

Naslund MJ, Strandberg JD, Coffey DS. 1988. The role of androgens and estrogens in the pathogenesis of experimental nonbacterial prostatitis. *J Urol* 140:1049-53.

Nelson WG, De Marzo AM, Isaacs WB. 2003. Prostate cancer. *N Engl J Med* 349(4)**:** 366-381.

Nickel JC, Roehrborn CG, O'leary MP, Bostwick DG, Somerville MC, Rittmaster RS. 2007. Examination of the relationship between symptoms of prostatitis and histological inflammation: baseline data from the REDUCE chemoprevention trial. *J Urol* 178:896-900.

Nickel JC, Roehrborn CG, O'Leary MP, Bostwick DG, Somerville MC, Rittmaster RS. 2008. The relationship between prostate inflammation and lower urinary tract symptoms: examination of baseline data from the REDUCE trial. *Eur Urol* 54(6):1379-84.

NTP. 2008. National Toxicology Program. Center for the Evaluation of Risks to Human Reproduction. NTP-CERHR monograph on the potential human reproductive and developmental effects of bisphenol A. NIH Publication No. 08-5994. Research Triangle Park, NC.

Orsi L, Delabre L, Monnereau A, Delval P, Berthou C, Fenaux P, Marit G, Soubeyran P, Huguet F, Milpied N, Leporrier M, Hemon D, Troussard X, Clavel J. 2009. Occupational exposure to pesticides and lymphoid neoplasms among men: results of a French case-control study. *Occup Environ Med* 66(5):291-298.

Paul H, Dunsire J, Hedley D. 1993. The Absorption, Distribution, Degradation and Excretion of (U-carbon 14) Triazine G 30027 in the Rat: Lab Project Number: 9523: 523-92: 153138. Unpublished study prepared by Inveresk Research International. 178 p. {OPPTS 870.7485}. MRID: 44713802

Rayner JL, Enoch RR, Wolf DC, Fenton SE. 2007. Atrazine-induced reproductive tract alterations after transplacental and/or lactational exposure in male Long-Evans rats. *Toxicol Appl Pharmacol* 218(3):238-48.

Robinette CL. 1988. Sex-hormone-induced inflammation and fibromuscular proliferation in the rat lateral prostate. *Prostate* 12:271-86.

Ross KM, Jones TL, Filipov NM. 2009. Disposition of the Herbicide 2-Chloro-4-(ethylamino)-6-(isopropylamino)-s-triazine (Atrazine) and Its Major Metabolites in Mice: A Liquid Chromatography/Mass Spectrometry Analysis of Urine, Plasma, and Tissue Levels. Drug Metabolism and Disposition. *Drug Metab Dispos* 37:776-786.

Ruder AM, Waters MA, Butler MA, Carreón T, Calvert GM, Davis-King KE, Schulte PA, Sanderson WT, Ward EM, Connally LB, Heineman EF, Mandel JS, Morton RF, Reding DJ, Rosenman KD, Talaska G. 2004. Gliomas and farm pesticide exposure in men: the upper midwest health study. *Arch Environ Health* 59(12):650-657.

Rull RP, Gunier R, Von Behren J, Hertz A, Crouse V, Buffler PA, Reynolds P. 2009. Residential proximity to agricultural pesticide applications and childhood acute lymphoblastic leukemia. *Environ Res* 109(7):891-899.

Rusiecki JA, De Roos A, Lee WJ, Dosemeci M, Lubin JH, Hoppin JA, Blair A, Alavanja MCR. 2004. Cancer incidence among pesticide applicators exposed to atrazine in the Agricultural Health Study. *J Natl Cancer Inst* 96:1375–1382.

Ryberg KR, Vecchia AV, Martin JD, Gilliom RJ. (2010). Trends in pesticide concentrations in urban streams in the United States, 1992–2008: U.S. Geological Survey Scientific Investigations Report 2010–5139, 101 p.

Schroeder JC, Olshan AF, Baric R, Dent GA, Weinberg CR, Yount B, Cerhan JR, Lynch CF, Schuman LM, Tolbert PE, Rothman N, Cantor KP, Blair A. 2001. Agricultural risk factors for t(14;18) subtypes of non-Hodgkin's lymphoma. *Epidemiology* 12(6):701-709.

Seethalakshmi L, Bala RS, Malhotra RK, Austin-Ritchie T, Miller-Graziano C, Menon M, Luber-Narod J. 1996. 17 beta-estradiol induced prostatitis in the rat is an autoimmune disease. *J Urol* 156:1838-42.

Simoneaux BJ. 1985. Metabolism of 14-C atrazine in orally dosed rats (general metabolism). Laboratory Study No. ABR-85104. Biochemistry Department, Agricultural Division, Ciba-Geigy Corp. MRID: 40431309.

Stanko JP, Enoch RR, Rayner JL, Davis CC, Wolf DC, Malarkey DE, Fenton SE. 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* 30(4):540-9.

Steiner GE, Stix U, Handisurya A, Willheim M, Haitel A, Reithmayr F. Paikl D, Ecker RC, Hrachowitz K, Kramer G, Lee C, Marberger M. 2003. Cytokine expression pattern in benign prostatic hyperplasia infiltrating T cells and impact of lymphocytic infiltration on cytokine mRNA profile in prostatic tissue. *Lab Invest* 83(8):1131-46.

Stoker TE, Robinette CL, Cooper RL. 1999. Maternal exposure to atrazine during lactation suppresses suckling-induced prolactin release and results in prostatitis in the adult offspring. *Toxicol Sci* 52(1):68-79.

Tangbanluekal L, Robinette CL. 1993. Prolactin mediates estradiol-induced inflammation in the lateral prostate of Wistar rats. *Endocrinology* 132:2407-16.

Thede B. 1987. Study of [Carbon 14]-atrazine dose/response relationship in the rat (general metabolism): Laboratory Study No. ABR-87087. Unpublished study prepared by Ciba-Geigy Corp. MRID: 40431305.

Thorpe N, Shirmohammadi A. 2005. Herbicides and nitrates in groundwater of Maryland and childhood cancers: a geographic information systems approach. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev* 23(2):261-278.

Timms BG, Mohs TJ, Didio LJ. 1994. Ductal budding and branching patterns in the developing prostate. *J Urol* 151(5):1427-32.

Timms BG, Howdeshell KL, Barton L, Bradley S, Richter CA, vom Saal FS. 2005. Estrogenic chemicals in plastic and oral contraceptives disrupt development of the fetal mouse prostate and urethra. *Proc Natl Acad Sci* U S A 102(19):7014-9.

Timms BG. 2008. Prostate development: a historical perspective. *Differentiation* 76(6):565-77.
US EPA. 2010. Framework for incorporating human epidemiologic & incident data in health risk assessment, January 7, 2010. Presented at FIFRA Scientific Advisory Panel Meeting, February 2-5, 2010.

US CDC. 2011. Breastfeeding Report Card—United States, 2011. http://www.cdc.gov/breastfeeding/data/reportcard.htm.

Van Leeuwen J, Waltner-Toews D, Abernathy T, Smit B, Shoukri M. 1999. Associations between stomach cancer incidence and drinking water contamination with atrazine and nitrate in Ontario (Canada) agroecosystems, 1987-1991. *Int J Epidemiol* 28(5):836-840.

WHO. 2010. Characterization and Application of Physiologically Based Pharmacokinetic Models in Risk Assessment. International Programme on Chemical Safety. IPCS Harmonization Project. http://www.who.int/ipcs/methods/harmonization/areas/pbpk_models.pdf).

Young H, Mills P, Riordan D, Cress R. 2005. Triazine herbicides and epithelial ovarian cancer risk in central California. *Occup Environ Med* 47(11):1148-1156.

Zahm SH, Weisenburger DD, Cantor KP, Holmes FF, Blair A. 1993. The role of the herbicide atrazine in the development of non-Hodgkin's lymphoma. Scand J Work Environ Health 19(2): 108-114.

APPENDIX 1

Summary of Epidemiologic Study Results Published since 2003 SAP				
Reference	**Cancer Endpoint and Sample Size**	**Exposure Assessment**	**Adjustment**	**Risk Ratio**
Nested Case-Control Studies (AHS)				
DeRoos 2003	Non-Hodgkin's Lymphoma 650 incident cases, 1933 controls (pooled from 3 studies)	self-reported use atrazine	Age, use of other pesticides, study site	OR 1.6 (95% CI 1.1, 2.5) in logistic regression; OR 1.5 (95% CI 1.0, 2.2)
Alavanja 2003	Prostate 566 cases, 54,766 controls	self-reported use atrazine	age, family history of prostate cancer	ever-use atrazine: OR 0.94 (95% CI 0.78-1.14) (OR per quintile LTED ~1.0, p-trend 0.34)
Alavanja 2004	Lung 300 cancer cases 57,284 pesticide applicators and 32,333 spouses;	self-reported use atrazine	n/a	no exposure-response trend observed triazines/atrazine; no results shown
Engel 2005	Breast 309 cases, 30,454 controls	self-reported use atrazine	age, race, state	OR 0.70 (95% CI 0.4, 1.2) among wives who directly apply
Lee 2007	Colorectal 305 cases, 56,813 controls	self-reported use atrazine	Age, smoking, state, total days of pesticide application among all enrolment applicators.	Colorectal: ever-use atrazine OR 0.9 (95% CI 0.70, 1.2); Colon OR 0.8 (95% CI 0.60, 1.1); Rectum OR 1.2 (95% CI 0.70,2.0)
Andreotti 2009	Pancreatic 93 incident pancreatic cancer cases, 82,503 cancer-free controls	self-reported use atrazine	age, smoking, diabetes, applicator type,5 most highly correlated pesticides	Ever vs. never atrazine use OR 0.70 (95% CI 0.4-1.2)
Dennis 2010	Cutaneous Melanoma 150 incident melanomas among 24,704 applicators	self-reported use atrazine	n/a	no exposure-response trend observed triazines/atrazine; no results shown
Cohort (AHS)				
Rusiecki 2004	Prostate 554 incident cancers	self-reported use atrazine	Age, sex, alcohol use, farm residence, smoking, education, family history of cancer, state, use of 10 highly correlated	Lifetime Exposure Days (LTED) quartiles vs. lowest quartile (1-20 days) 21-56 days OR 0.89 (95% CI 0.66, 1.21), 57-178.5 days OR 0.75 (95% CI

	<u>Lung</u> 118 incident cancers		pesticides	0.56, 1.02), and >178.5 days OR 0.88 (95% CI 0.63, 1.23), p=0.26 for linear trend LTED same quartiles as above vs. lowest quartile OR 0.87 (95% CI .4, 1.87), OR 1.13 (95% CI .056, 2.29), OR 1.91 (0.93, 3.94), p for trend =0.08, for IWED p for trend = 0.19
	<u>Bladder</u> 47 incident cases			LTED same quartiles as above vs. lowest quartile OR 2.25 (0.67, 7.62), OR 1.04 (0.27, 4.05), OR 3.06 (0.86, 10.81), p for trend 0.18, ns for IWED Ns relations for oral cavity, esophagus, colon, rectum, pancreas, melanoma, kidney, NHL, multiple myeloma, leukemia

Alavanja 2005	<u>Prostate</u> 1087 incident cases	Self-reported use atrazine	Age, calendar year	Observed/expected (O/E) = 1.26 (95% CI 1.18, 1.33) in private applicators and O/E =1.37 (95% CI 0.98, 1.86) in commercial applicators
	<u>Ovary</u> 8 incident cases in private applicators			O/E = 2.97 (95% CI 1.28, 5.85) Ns relations with buccal cavity, digestive system, respiratory system, soft tissue, melanomas, breast, urinary system brain, thyroid, leukemias and lymphomas
	<u>Non-Hodgkin's Lympoma</u> 114 incident cases in			O/E = 1.02 (95% CI 0.84, 1.22) for private applicators; O/E = 0.96 (95% CI 0.35, 2.10) in

	private applicators, 6 incident cases in commercial applicators			commercial applicators
Koutros 2010	<u>Prostate</u> 1719 incident cases in private applicators, 73 in commercial applicators <u>Ovary</u> 9 incident cases in private applicators, 0 in commercial applicators	Self-reported use atrazine	Age, year	Standardized Incidence Ratio (SIR) 1.19 (95% CI 1.14, 1.25) in private applicators, SIR 1.28 (95% CI 1.00, 1.61) in commercial applicators for ovarian cancer SIR 2.45 (95% CI 1.12, 4.,65) in private applicators

Beane Freeman 2011	<u>Prostate</u> 1297 incident cancers <u>Lung</u> 275 incident cancers <u>Thyroid</u> 29 incident cancers in men <u>Breast cancer</u> 36 incident cancers in women <u>Ovarian cancer</u> 9 (4 exposed) incident cancers <u>Non-Hodgkin's Lymphoma</u> 152 incident cases	self-reported use atrazine with updated exposure algorithm including accounting for protective factors and spray application methods	Age, race, gender, smoking, alcohol use, education, state, family history of specific cancers, applicator type (private or commercial), and every use of other pesticides highly correlated with atrazine	No association No association LTED upper vs. lowest quartile OR 2.32 (95% CI 0.66, 8.22), intensity weighted exposure days (IWED) OR 4.84, (95% CI 1.31, 17.93), ns linear trend but linear trend p=0.08 for IWED, non-monotonic exposure-response trend Ever vs. never use OR 1.14 (95% CI 0.47, 2.50), upper vs. lower median OR Ever vs. never use OR 2.91 (95% CI 0.56, 13.6) No significant association or linear trend with LTED or IWED
Hospital-based case-control				

Orsi 2009	Non-Hodgkin's lymphoma 244 cases in men Hodgkin's lymphoma 87 cases in men Multiple myeloma 56 cases in men	Self-reported triazine herbicides	Age, center	OR 1.9 (95% CI 0.9, 3.8) OR 3.2 (95% CI 0.9, 10.9) OR 1.7 (95% 0.5, 5.9)

Population-based case-control				
Young 2005	Ovarian cancer 256 incident cases, 1122 controls	Self-reported triazine herbicides	Age, family history of breast and/or ovarian cancer, use of oral contraceptives or menopausal hormone therapy, pregnancy and breastfeeding history, race/ethnicity	Ever vs. never occupational use of triazines OR 1.34 (95% CI 0.77, 2.30), non-significant, non-monotonic linear trend p=0.22; ever vs. never use of atrazine OR 0.76 (95% CI 0.16, 3.55)

SUMMARY BY CANCER TYPE

Prostate

Reference	Study Design and Sample Size	Exposure Assessment	Adjustment	Risk Ratio
Alavanja 2003	Nested Case-Control (AHS), 566 cases	self-reported use atrazine	age, family history of prostate cancer	ever-use atrazine: OR 0.94 (95% CI 0.78-1.14) (OR per quintile LTED ~1.0, p-trend 0.34)
Rusiecki 2004	Cohort (AHS) 554 cases	Self-reported use of atrazine	Age, sex, alcohol use, farm residence, smoking, education, family history of cancer, state, use of 10 highly correlated pesticides	Lifetime Exposure Days (LTED) quartiles vs. lowest quartile (1-20 days) 21-56 days OR 0.89 (95% CI 0.66, 1.21), 57-178.5 days OR 0.75 (95% CI 0.56, 1.02), and >178.5 days OR 0.88 (95% CI 0.63, 1.23), p=0.26 for linear trend
Alavanja 2005	Cohort (AHS) 1087 incident cases	Self-reported atrazine use	Age, calendar year	Observed/expected (O/E) = 1.26 (95% CI 1.18, 1.33) in private applicators and O/E =1.37 (95% CI 0.98, 1.86) in commercial applicators
Koutros 2010	Cohort (AHS) 1719 incident	Self-reported atrazine use	Age, year	Standardized Incidence Ratio (SIR) 1.19 (95% CI 1.14, 1.25) in private applicators,

	cases in private applicators, 73 in commercial applicators			SIR 1.28 (95% CI 1.00, 1.61) in commercial applicators
Beane Freeman 2011	Cohort (AHS) 1297 incident cancers	Self-reported atrazine with updated protective and application factors	Age, race, gender, smoking, alcohol use, education, state, family history of specific cancers, applicator type (private or commercial), and every use of other pesticides highly correlated with atrazine	No significant association, no significant linear trend

Breast

Reference	Study Design and Sample Size	Exposure Assessment	Adjustment	Risk Ratio
Engel 2005	Nested case-control (AHS) 309 cases, 30,454 controls	self-reported use atrazine	age, race, state	OR 0.70 (95% CI 0.4, 1.2) among wives who directly apply
Beane Freeman 2011	Cohort 36 incident cases	self-reported use atrazine with updated exposure algorithm including accounting for protective factors and spray application methods	Age, race, gender, smoking, alcohol use, education, state, family history of specific cancers, applicator type (private or commercial), and every use of other pesticides highly correlated with atrazine	Ever vs. never use OR 1.14 (95% CI 0.47, 2.50), upper vs. lower median OR

Ovary

Reference	Study Design and Sample Size	Exposure Assessment	Adjustment	Risk Ratio
Alavanja 2005	Cohort (AHS) 8 incident cases	Self-reported use atrazine	Age, calendar year	O/E = 2.97 (95% CI 1.28, 5.85)

	in private applicators			
Beane Freeman 2011	Cohort (AHS) 9 incident cancer	Self-reported atrazine with updated protective and application factors	Age, race, gender, smoking, alcohol use, education, state, family history of specific cancers, applicator type (private or commercial), and every use of other pesticides highly correlated with atrazine	Ever vs. never use OR 2.91 (95% CI 0.56, 13.6)
Koutros 2010	9 incident cases in private applicators, 0 in commercial applicators	Self-reported use atrazine	Age, year	Standardized Incidence Ratio (SIR) 1.19 (95% CI 1.14, 1.25) in private applicators, SIR 1.28 (95% CI 1.00, 1.61) in commercial applicators for ovarian cancer SIR 2.45 (95% CI 1.12, 4.,65) in private applicators
Young 2005	Population-based case-control 256 incident cases, 1122 controls	Self-reported triazine herbicides	Age, family history of breast and/or ovarian cancer, use of oral contraceptives or menopausal hormone therapy, pregnancy and breastfeeding history, race/ethnicity	Ever vs. never occupational use of triazines OR 1.34 (95% CI 0.77, 2.30), non-significant, non-monotonic linear trend p=-.22; ever vs. never use of atrazine OR 0.76 (95% CI 0.16, 3.55)

Thyroid

Reference	Study Design and Sample Size	Exposure Assessment	Adjustment	Risk Ratio
Beane Freeman 2011	Cohort (AHS) 29 incident cancers in men	Self-reported atrazine with updated protective and application factors	Age, race, gender, smoking, alcohol use, education, state, family history of specific cancers, applicator type (private or	LTED upper vs. lowest quartile OR 2.32 (95% CI 0.66, 8.22), intensity weighted exposure days (IWED) OR 4.84, (95% CI 1.31, 17.93), ns linear trend but linear trend p=0.08 for IWED, non-monotonic exposure-response trend

			commercial), and every use of other pesticides highly correlated with atrazine	

Non-Hodgkin's Lymphoma

Reference	Study Design and Sample Size	Exposure Assessment	Adjustment	Risk Ratio
DeRoos 2003	Nested case-control 650 incident cases; 1933 controls	self-reported use atrazine	Age, use of other pesticides, study site	OR 1.6 (95% CI 1.1, 2.5) in logistic regression; OR 1.5 (95% CI 1.0, 2.2)
Alavanja 2005	Cohort 114 incident cases in private applicators, 6 incident cases in commercial applicators	self-reported use atrazine	Age, calendar year	O/E = 1.02 (95% CI 0.84, 1.22) for private applicators; O/E = 0.96 (95% CI 0.35, 2.10) in commercial applicator
Orsi 2009	Hospital-based case-control 244 cases in men	Self-reported triazine herbicides	Age, center	OR 1.9 (95% CI 0.9, 3.8)
Beane Freeman 2011	Cohort 152 incident cases	Self-reported atrazine with updated protective and application factors	Age, race, gender, smoking, alcohol use, education, state, family history of specific cancers, applicator type (private or commercial), and every use of other pesticides highly correlated with atrazine	No significant association or linear trend with LTED or IWED