SAP Report No. 2000-03
September 15, 2000

REPORT

FIFRA Scientific Advisory Panel Meeting, June 6-7, 2000, held at the Sheraton Crystal City Hotel, Arlington, Virginia

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

Session I - Consultation: National Drinking Water Survey Design for Assessing Chronic Exposure

NOTICE

This report has been written as part of the activities of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP). This report has not been reviewed for approval by the United States Environmental Protection Agency (Agency) and, hence, the contents of this report 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 was established under the provisions of FIFRA, as amended by the Food Quality Protection Act (FQPA) of 1996, to provide 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 EPA, Office of Pesticide Programs (OPP) and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. Food Quality Protection Act 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 Larry Dorsey, SAP Executive Secretary, via e-mail at dorsey.larry@.epa.gov.

TABLE OF CONTENTS

Session I: Consultation: National Drinking Water Survey Design for Assessing Chronic Exposure

SAP Report No. 2000-03A, September 15, 2000

REPORT:
FIFRA Scientific Advisory Panel Meeting,
June 6, 2000, held at the Sheraton Crystal City Hotel, Arlington, Virginia

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

Consultation: National Drinking Water Survey Design for Assessing Chronic Exposure

Mr. Paul Lewis
Designated Federal Official
FIFRA Scientific Advisory Panel
Date:______________________

Herbert Needleman, M.D.
FIFRA SAP Session Chair
FIFRA Scientific Advisory Panel
Date:______________________

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
June 6, 2000

SESSION I - Consultation: National Drinking Water Survey Design for Assessing Chronic Exposure

PARTICIPANTS

FIFRA SAP Session Chair
Herbert Needleman, M.D., Professor of Psychiatry and Pediatrics, University of Pittsburgh School of Medicine, Pittsburgh, PA

FIFRA Scientific Advisory Panel
Mary Anna Thrall, D.V.M., Professor, Department of Pathology, College of Veterinary Medicine & Biomedical Sciences, Colorado State University, Fort Collins, CO

FQPA Science Review Board Members
Paul Capel, Ph.D., USGS, Department of Civil Engineering, University of Minnesota, MN
Scott Ferson, Ph.D., Applied Biomathematics, Setauket, NY
George Hallberg, Ph.D., The Cadmus Group, Inc., Waltham, MA
Steve Heeringa, Ph.D., Institute for Social Research, University of Michigan, Ann Arbor, MI
Richard Hooper, Ph.D., USGS, Northborough, MA
Robert Kellogg, Ph.D., National Resource Conservation Service, USDA, Washington, DC
Mr. Robert O' Brien, Battelle, Pacific Northwest Labs, Richland, WA
Kenneth M. Portier, Ph.D., Associate Professor, Department of Statistics, Institute of Food and Agricultural Sciences and Natural Resources, University of Florida, Gainesville, FL
R. Peter Richards, Ph.D., Director, Water Quality Laboratory, Heidelberg College, Tiffin, OH
Harold Van Es, Ph.D., Cornell University, Department of Crop and Soil Sciences, Ithaca, NY

Designated Federal Official
Mr. Paul Lewis, FIFRA Scientific Advisory Panel, Office of Science Coordination and Policy, Office of Prevention, Pesticides and Toxic Substances, Environmental Protection Agency, Washington, DC

PUBLIC COMMENTERS

Oral statements were made by:
Ray McAllister, Ph.D. on behalf of the American Crop Protection Association
Peter Coody, Ph.D. on behalf of the American Crop Protection Association
Russell Jones, Ph.D. on behalf of Aventis CropScience
Ms. Jane Houlihan on behalf of the Environmental Working Group
Mr. David Marker on behalf of Westat, Inc.

Written statements were received from:
American Crop Protection Association
Environmental Working Group

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) has completed its review of the set of scientific issues being considered by the Agency pertaining to a consultation on a national drinking water survey design for assessing chronic exposure. Advance notice of the meeting was published in the *Federal Register* on May 17, 2000. The review was conducted in an open Panel meeting held in Arlington, Virginia, on June 6, 2000. The meeting was chaired by Herbert Needleman, M.D. Mr. Paul Lewis served as the Designated Federal Official.

Under the Food Quality and Protection Act of 1996, drinking water is considered in aggregate exposure assessments for pesticide tolerance reassessments. Since targeted monitoring data are needed for refined exposure assessments, the EPA Office of Pesticide Programs (OPP) has proposed a national drinking water design framework for assessing annual average pesticide concentrations in surface waters used as drinking water. Details of survey design issues and options were presented on OPP's proposed design framework and on an independently proposed framework.

Mah Shamin, Ph.D., and Ms. Elizabeth Behl (EPA, Office of Pesticide Programs) opened the session by providing background on the topics to be discussed. James Cowles, Ph.D. (EPA, Office of Pesticide Programs) and Ian Kennedy, Ph.D. (EPA, Office of Pesticide Programs) discussed the goals and objectives of drinking water monitoring. Norman Birchfield, Ph.D. (EPA, Office of Pesticide Programs) and Rodolfo Pisigan, Ph.D. (EPA, Office of Pesticide Programs) provided an overview of key issues and options for a proposed national drinking water survey design. James Hetrick, Ph.D. (EPA, Office of Pesticide Programs) and R. David Jones, Ph.D. (EPA, Office of Pesticide Programs) presented the EPA, OPP, Environmental Fate and Effects Division's suggested approach to a national drinking water design framework.

CHARGE

1.) EFED is recommending the following factors to identify target pesticides for inclusion in the study: extent of use area, environmental fate properties, and risk.
Is this a reasonable strategy for selecting chemicals for the survey design?

- Would some other approach better serve the long-term objective of developing and testing predictive models?

2.) EFED is recommending that pesticide use area be the primary domain. In order to better define exposure levels to specific human populations and to better understand the processes affecting pesticide concentrations in surface source drinking water, it may be useful to use smaller domains. Such domains might be spatial or based on vulnerability criteria.

- Should domains other than a pesticide's use area be used (regional domains, for example)?
- What are the panel's views on minimally acceptable data quality standards if the study had a larger number of domains.
- Are there any suggestions for optimizing survey design to maximize the number of target pesticides represented while minimizing the total number of sampled CWSs without sacrificing data quality?
- Could preferential selection of CWSs with source watersheds containing uses for multiple pesticides be incorporated into the design?

3.) EFED is considering stratifying the CWSs by vulnerability. The estimation of vulnerability would attempt to identify, for each pesticide and CWS, the likelihood of contamination. EFED would then randomly select CWSs from the vulnerability strata for inclusion in the survey.

- How should we combine factors into an estimate of site vulnerability?
- Would additional or different strata increase the usefulness of the data in addressing the objectives of the monitoring study?
- What are the relative merits of using vulnerability-based as opposed to geographically-based stratification methods?

4.) EFED is proposing to use a geographic information systems (GIS) watershed characterization tool for identifying CWSs in the domains and strata.

- Does the SAP agree with the usefulness of this tool for these purposes?
- We are aware of limitations in the accuracy of the data used in the site selection tool. Is the panel aware of better ways of handling the accuracy and precision limitations of this data?
- Is the Panel aware of alternative data sources that may not have these limitations?

5.) EFED recommends collecting raw water and finished water samples with reactive analysis of the finished water if pesticides are detected in the raw water. Finished water more directly reflects drinking water exposure, while raw water better reflects watershed and usage characteristics allowing better risk management decisions.

- Given the analytical costs for paired raw and finished water samples and the lack of comprehensive information on water treatment effects on the removal and transformation of most pesticides, does the panel have alternative recommendations on assessing raw and finished water?

6.) Preliminary analyses by EFED indicates pesticide fate properties, sampling frequency, and hydrologic residence time impact the accuracy of calculated annual mean concentrations.

- Does the SAP have recommendations on balancing the characteristics of the pesticides and the water body being sampled with the number of samples needed to adequately measure the annual mean at different CWSs?

7.) Annual mean pesticide concentrations occurring at any CWS vary from year to year. A multi-year study would help to quantify year-to-year variability but is more costly.

- Does the SAP have any suggestions for assessing annual variability given the financial constraints of the survey?
- Would drawing out the survey over three years (with the same number of samples per CWS) improve it?

8.) EFED recognizes that pesticide concentrations in drinking water are dependent on factors including watershed characteristics, pesticide use, pesticide fate properties, surface water hydrology, and water treatment processes. Interpretation of the monitoring data will be dependent on the collection of such related ancillary data.

- What types of ancillary information does the SAP believe would assist in the interpretation of the monitoring data, and application of the data to model development and validation?

9.) We have defined our population as CWSs with pesticide use in their watersheds, based on the assumption that runoff and near field spray drift are the major routes of loading to the water supply.

- Is this a reasonable assumption and should we monitor facilities that do not have pesticide use in the watershed?

10.) One option proposed in the design framework is a census of the facilities serving the largest cities. These facilities are believed to have little agriculture in their watersheds, but reflect the drinking water for a large percentage of the population.

- Does the SAP believe sampling CWSs serving very large populations is useful even if they are expected to be of low vulnerability?

11.) Modeling should allow the Agency to reduce the extent of future drinking water surveys, and help to better identify areas at higher risk.

- Are there any changes to the design which will better allow the survey to support model development and testing?

GENERAL COMMENTS

The SAP thanked the Agency staff for their excellent presentations and the obvious level of effort that had gone into the proposal to date. The SAP noted that this was obviously a work in progress, and that the design needed considerable refinement. Because many issues require further clarification, the SAP noted that they would perhaps provide more questions than answers. The SAP supported the general goals of the design to: 1) obtain accurate estimates of chronic pesticide exposure in public drinking water systems using surface water sources, and 2) to collect information to develop and/or improve predictive models. However, the presentations implied that many other goals and objectives were being considered in the design. Hence, the details of the design remain too general for a full technical assessment of the approach. The likelihood of success will be dependent upon decisions that have not yet been explicitly made.

It was also unclear how this design may or may not relate or need to relate to considerations of studies to address acute exposure. The SAP was concerned that the level of funding being discussed may not be adequate. Also, there was strong consensus that the study design should not be dictated by an expected level of funding. An appropriate design should be considered and then a cost estimate would be constructed. The SAP noted that many of the sub-objectives need to be more clearly worked out so that the details of the design could be adequately addressed and reviewed.

The SAP also expressed concern that it was not clear how past studies and information had been used to guide the design. It was not evident if the extensive sampling programs of USGS and EPA had been used as "pilot" data to reach some of the design conclusions. Also, there was some concerns about details of the sample universe and some practical sampling issues. For example, it was noted that there were approximately 10,700 Community Water Systems (CWSs) using surface water that might constitute the sample population. The SAP noted that, from their knowledge, this figure must include purchased water systems that shouldn't be part of the sample. The sample universe is more likely about 5,500 system that use surface water as their source. This may make a difference in how the design is viewed. Also, is a CWS considered a single sample site? Many CWSs have multiple source intakes. Sampling the finished drinking water is typically done at the entry point to the distribution system; many systems have many entry points that may need to be sampled to characterize the finished water. Alternatively, would sampling from taps within the system be used to integrate samples. Such issues will affect the design and cost estimates.

The SAP noted this was a good start and that the Agency should refine its objectives and then develop a proposal, with clear goals and budgets, for further review. Given the uncertainty with many aspects of the design, the SAP also suggested that before proceeding with a National survey, a pilot survey on a local regional scale should be carried out to test estimation techniques, stratification issues, individual CWS sampling issues and the validity of the many assumptions used in building the study design. Also, it was noted that not all the information desired had to be collected across the entire survey. Some subsets of the samples in the study could address various details to reduce design and cost problems.

Other general comments are summarized by topic. These comments are followed by responses to the specific charge to the Panel.

Study Design and Funding

The SAP agreed with the Agency that the general goals of monitoring need to be two-fold: 1) obtain accurate estimates of exposure for as many pesticides as economic constraints will allow, and 2) collect the information so as to maximize use of the data to construct predictive models. One of the objectives of the monitoring study is to try to collect enough information on several pesticides in a single effort so that continuous monitoring is not required to estimate future exposure. The only way this objective can be met is if these data can be used to construct predictive models which could subsequently be applied to predict exposure in the future as land use patterns change, pesticide use patterns change, new pesticides are introduced, and precipitation changes.

However, the level of funding specified for this effort--7-10 million dollars--seems inadequate to meet the stated goals. At this level of funding, exposure estimates for only a handful of pesticides could be made with the precision needed. More importantly, this level of coverage would not be adequate to produce a predictive model that would estimate exposure reliable enough to negate the need for additional monitoring over time. The proposal by the American Crop Protection Association (ACPA) calls for a single-year study, which will not capture significant changes in weather or year-to-year variability in land management practices and pesticide use by agricultural producers. The ACPA statistical design, on which the cost estimate of 7-10 million dollars was initially based, seems flawed in the degree to which it assumes CWSs (and the watersheds providing the water for CWSs) to be homogenous. An appropriate statistical design must allow for the variability from year to year and the variability among CWSs relative to changes in pesticide use and pesticide fate and transport within each watershed. To collect sufficient monitoring data, in a one-shot sampling program that will be adequate for generating reliable prediction models so that annual monitoring would not be needed, would require a minimum investment of something on the order of 50 million dollars spread over a three year period. Even then, there would be a need to continue monitoring at some level to verify that the model predictions were on track, and to provide information needed to upgrade the models to predict exposure levels for new pesticides. A single-year monitoring effort funded at the level of 7-10 million can at best produce exposure estimates for one point in time for a handful of pesticides, requiring a great deal of additional investment to address the remaining pesticides and to address changes in exposure over time.

Statistical Considerations

The SAP understands that sophisticated analyses were used to determine the sample size required to achieve a 95% confidence about the 95th percentile. These methods are used to set the data quality standard but focusing on a single percentile seems unnatural and unnecessary, similar to describing a human's body shape by specifying the size of its nose. One SAP member suggested that the Agency consider using Kolmogorov-Smirnov confidence intervals for distributions instead. These confidence intervals express the uncertainty arising from limited sampling for the distribution as a whole, rather than for just a single percentile. The method is distribution-free, so that it doesn't make any obviously untenable assumptions. The method is known to be somewhat conservative, although its conservativism may not be appropriate in this regulatory assessment context.

The Kolmogorov-Smirnov confidence limits are classically expressed as intervals on the probability, and there may be some issue in translating between uncertainty about p and uncertainty about the x-value. There might also be some issue—or perhaps some economy—in the translation from the underlying distribution of annual mean concentrations for CWSs to the 10-year extreme distribution. Perhaps the literature on extreme value theory should be consulted.

Another issue that seems insufficiently considered in designing monitoring protocols is measurement error. Although laboratory measurement error is rarely negligible, it is often ignored in subsequent statistical analysis and risk assessments, which are commonly preoccupied with sampling error. One Panel member recommended that the monitoring protocol insist that measurement error be (1) recorded, (2) reported, and (3) propagated. Primary data measurements such as chemical determinations should be recorded with error intervals that summarize the precision of laboratory protocols that produced them. These intervals should be reported in all derived data sets and summaries based on the data. Finally, any calculations that make use of the measurements should also propagate measurement errors, at least with some simple bounding analysis. Although propagating the measurement error through calculations is obviously more trouble than simply ignoring it (which is what using only best estimates does), doing so expresses the reliability of any conclusions that are based on the measurements.

There appears to be a great deal of confusion between domains and strata in the presentations and the background material. It appears that domains refer to populations of interest and that strata refer to survey design units that can be aggregated to domains of interest. Clarification should be provided in future documents clearly specifying the domains of interest and design strata and how they interrelate.

The methodology for determining the number of sites per pesticide use area only ensures that with a given probability that the observed maximum annual average CWS pesticide concentration will exceed the true 95th percentile. This does not ensure that the precision of the point estimates of interest will be suitable for inferential purposes of this study.

Sample Numbers and Sampling

The assumption implicit in the Agency background document is that there is a one-time study where a number is measured and the job is done. One SAP member noted this is technically indefensible. The level of uncertainty in temporal patterns, for example, preclude the choice of a single number (be it 10 or 14 samples per year) as the appropriate number of samples to collect. It was suggested that the problem be approached with an explicit realization that multiple "rounds" of sampling will be necessary and that an adaptive design may be necessary.

It was suggested that technological advances that permit sample compositing might be looked at, as a way to integrate temporal sampling. Various problems were noted (analyte degradation, etc.) but it was agreed this might be reviewed for portions of the study.

Observations On Occurrence

Some of the highest concentrations seen in monitoring programs do not appear to be associated with storm runoff. Rather, they appear to represent spills or dumps of pesticides directly into the stream or river. They are not frequent, fortunately, but one such spill could have a substantial effect on the annual mean concentration. There isn't any realistic sampling strategy that has a good chance of capturing these events, but they must be acknowledged.

DETAILED RESPONSE TO THE CHARGE

The specific issues to be addressed by the Panel are keyed to the Agency's background document "FIFRA Scientific Advisory Panel Briefing Document for a Consultation on Monitoring Strategies for Pesticides in Surface Derived Drinking Water", dated May 8, 2000, and are presented as follows:

1.) EFED is recommending the following factors to identify target pesticides for inclusion in the study: extent of use area, environmental fate properties, and risk.

- **Is this a reasonable strategy for selecting chemicals for the survey design?**
- **Would some other approach better serve the long-term objective of developing and testing predictive models?**

The three factors, extent of use area, environmental fate properties, and risk, should indeed be the primary factors in the process for selecting chemicals in a national drinking water survey design. These are factors that have been reviewed and discussed for years as key elements for such designs. None of the individual selection criteria provides a logical or completely satisfactory set of target pesticides. This monitoring study has multiple objectives, each requiring a different emphasis on pesticide selection factors. Therefore, the combination of criteria should be used, although not necessarily equally weighted.

Extent of use area is appropriate for a national survey of this scope that will focus on multiple pesticides. A national focus will also help insure that a range of climatic-hydrogeologic conditions is encompassed for model development. Generally, high risk pesticides should be included. The Agency stated that the list of pesticides requiring exposure estimates have been reduced using screening models (pesticides that are not expected to occur in water) and using extant monitoring data. Candidates high on the list included pesticides where the "risk cup" was partially (or nearly) full based only on exposure through food. This is the appropriate priority for meeting the goals of the Food Quality Protection Act, but it cannot be expected to provide enough variability in fate and transport characteristics to construct prediction models that are robust enough to make estimates of pesticide exposure for pesticides not included in the survey or for new pesticides. The SAP raised the question if non-human, ecological risk should also be considered, particularly if raw water occurrence was also a target for the survey, in addition to finished drinking water.

For evaluation of model performance, one needs to make sure that a reasonable range of pesticide fate properties are represented (as well as a range of environmental conditions). Additional pesticides might need to be added because of their fate and transport characteristics. Without a "representative" set of pesticides relative to fate and transport characteristics, reliable predictive models cannot be estimated, and additional monitoring might be required to make exposure estimates for the remaining pesticides. Therefore, the multiple objectives of this study require some flexibility in the selection procedures.

Also, another practical (and economic) factor that should be considered is what pesticides and transformation products might be covered by multi-analyte methods. Since cost containment may be a limiting factor in this monitoring study, it is important to maximize the information that can be obtained from efficient chemical analyses. However, this should not be the primary factor in selecting the target pesticide. Also, pesticide degradation/transformation products have been mentioned to be included. The purpose of including them should be carefully spelled out.

In addition, results from previous monitoring studies (USGS-NAWQA, etc.) should be consulted to identify "problem" chemicals that need to be included. Inclusion of chemicals that pose low probabilities of detection might be avoided, but this must be done with care. Advancing model development and testing the actual site selection protocol will be a critical component. A range of high to low use, a range of hydrogeologic, and fate and transport conditions must be covered in the site selection process.

2.) EFED is recommending that pesticide use area be the primary domain. In order to better define exposure levels to specific human populations and to better understand the processes affecting pesticide concentrations in surface source drinking water, it may be useful to use smaller domains. Such domains might be spatial or based on vulnerability criteria.

- **Should domains other than a pesticide's use area be used (regional domains, for example)?**
- **What are the Panel's views on minimally acceptable data quality standards if the study had a larger number of domains.**
- **Are there any suggestions for optimizing survey design to maximize the number of target pesticides represented while minimizing the total number of sampled CWSs without sacrificing data quality?**
- **Could preferential selection of CWSs with source watersheds containing uses for multiple pesticides be incorporated into the design?**

Many of these questions could be better answered after the objectives and some design ideas have been better defined. Even though the Panel agreed that additional analysis is needed, the SAP generally concurred with the Agency that the primary domain should be the pesticide use area. With the limited sampling that is being proposed, this seems the best way that reliable estimates of exposure for the selected pesticides can be made. Strictly geographic or spatial domains seem difficult to justify given the objectives. A set of regional domains can be applied, but only if the sampling rate were much higher. It could be advantageous to create subdomains of urban and rural environments, since the exposure for these two groups could be very different. Creation of subdomains, however, requires more sampling and must be carefully evaluated and justified.

In producing the pesticide use domains, however, better pesticide use data than that proposed in the study should be used if possible. The Agency has access to confidential use statistics that might produce estimates of use for recent years and sometimes for sub-state areas. While these data cannot be made public, it might be used to generate more precise and up-to-date pesticide use areas without endangering its confidentiality. Also, recent data collected by USDA should be used in conjunction with the Agency database. A further recommendation was averaging use over the most recent three years for which data are available to capture year-to-year variation in pesticide use (which in turn reflects year-to-year variation in crops planted and pest problems that are often linked to weather).

The spatial link between pesticide use domain (on the spatial level) and the actual sampling location (the public water system) is the watershed or source water area. This link should be used to optimize sample selection and design. Rather than using simple pesticide use (mass) as a criterion, mass normalized to area (mass/area) might be better to use as the basis for identifying and classifying the watersheds to be sampled. People drink water from a particular watershed, they do not drink average regional water. If this is not in the initial defining area, then the data have lost a considerable amount of utility. The use of the GIS tools discussed should allow these logistics to be addressed. Other domains probably are not as useful as being primary domains. The levels of pesticides in drinking water can be estimated for all or most other domains by aggregating the data based on pesticide use and watersheds.

This survey design seems to be about characterizing a distribution of distributions – or at least about characterizing the distribution of mean values of local site concentration distributions. The design seems to focus strongly on the larger distribution, and on a concern for characterizing the upper percentiles of this distribution. Choosing as a data quality criterion the goal to include the 95th percentile of the parent distribution with 95% confidence (and indeed, in each domain) places strong demands on the statistical design, but the choice of this specific criterion is not discussed. What is so magical about the 95th percentile? Why not be satisfied with the 95th percentile with lesser confidence? What should the relationship be between this data quality standard and risk-based standards such as MCLs?

One important thing that is lost in all of this is the issue of the sampling needed to adequately characterize the mean value of the local distribution – the individual data values which make up the overall distribution. Variability determines the number of samples needed to characterize the mean with a given level of confidence. How confident do we need to be about the individual mean values? There are statements in the Agency background document that indicate awareness that sampling needs will be greater for flowing-water systems than for reservoirs and other static-water systems, but no detailed investigation of this issue seems to have been made yet. One SAP member noted it would be good to see more attention paid to the data quality of the individual local mean concentrations and less focus on pinning down the upper tail of the larger distribution.

Some concern was expressed that question 2 seems contrary to the starting point of the survey. Should the amount of money spent be specified first, and the survey design and data quality be determined from that, or should the design and data quality be determined for the needs of the study at all necessary domains, the survey designed accordingly, then the cost established? This will lead to a more expensive program, but will provide the data that is really needed to address the questions. In reality, the design/cost/data quality should be conducted with an iterative approach until the optimum survey is designed. Starting with a predetermined cost is not the scientifically sound way of approaching this survey.

Several of the supporting documents describe stratification as a technique used to form relatively homogeneous groups, to reduce variance and increase confidence in results of a sampling program, for a given number of samples. However, the stratification schemes discussed in the background documents do not appear to fill this role. It is difficult to see how ACPA's suggestion to create five domains, one national and four regional, can identify or benefit from homogeneity. The suggestion of both ACPA and the Agency to stratify on vulnerability makes sense, because it can help to assure that more attention is paid to CWSs most likely to have relatively high mean concentrations, which is the part of the distribution that the Agency is most interested in. But again, it is not clear that this stratification will produce homogeneous groups, or increase statistical efficiency. More attention needs to be paid to stratification as a means of variance reduction, and to the complex and sometimes ambiguous relationship between sub-domains and strata.

Regarding the preferential selection of CWSs with multiple pesticides, in the Agency background document, the Agency reported that a full suite of analytes would be sampled at each CWS in the sample design, but that the design would be tailored to produce reliable estimates only for the targeted pesticides. The notion behind this question is that if just the right CWSs were selected, the number of pesticides with reliable estimates of exposure could be increased. While it may be possible to weigh the CWSs somewhat, it is critical that the design remain probabilistic. In the end, the CWSs chosen must represent the set not chosen, and there is the danger that with preferential selection you might draw a set of CWSs that are too unique. Instead of this approach, perhaps after the sample is selected it could be examined for cases where a few more samples in a pesticide use area would produce reliable estimates for that pesticide, and a second sample drawn to add the few additional CWSs needed to extend the list of pesticides for which reliable estimates could be obtained.

3.) EFED is considering stratifying the CWSs by vulnerability. The estimation of vulnerability would attempt to identify, for each pesticide and CWS, the likelihood of contamination. EFED would then randomly select CWSs from the vulnerability strata for inclusion in the survey.

- **How should we combine factors into an estimate of site vulnerability?**
- **Would additional or different strata increase the usefulness of the data in addressing the objectives of the monitoring study?**
- **What are the relative merits of using vulnerability-based as opposed to geographically-based stratification methods?**

The multiple pesticides of interest, the variability of hydrological, climatological, water body type features of individual sites present a very complex problem in stratification. Hence, this response has multiple components. Optimal stratification for a sample of CWS is relatively straightforward for a single analysis problem (a single pesticide, a single statistic of interest); however, the multiple objectives of this study preclude a choice of a single optimum for all questions of interest. For this reason, it is advisable that the Agency establish a prioritization of pesticides to study in depth based on grounds of scientific importance in the assessment of health risks. The SAP recognizes that the Agency, industry, and the public have concern over potentially hundreds of different pesticide compounds, but to achieve useful scientific and regulatory data from a study of this scope, hard choices over priorities must be made.

The definition of domains (collection of strata for analysis) that correspond to pesticide use areas and the differential sampling of CWSs based on a measure of the potential concentration in the water to be tested ("vulnerability") is a good stratification choice for the single pesticide. If a short list of priority pesticides can be established and the statistic of primary interest is the 95th percentile of the annual mean concentration for CWSs, then the use of strata defined by "vulnerability" or modeled predictions of concentrations of pesticides is a more optimal strategy compared to simply sampling at random from the CWSs within each use area domain. Optimal definition of strata boundaries and allocation of the total sample size to the defined strata will depend on the shape of the distribution of CWS annual mean concentrations including the proportion of CWSs that test below the LOD for the compound and the distributional skewness and thickness of its upper tail. Since these distributions will vary from one pesticide to another and possibly from one geographic region to the next, care should be taken to not rely totally on *a priori* predictions of the form of these distributions. It is important to emphasize that while oversampling strata of CWSs of expected higher concentration of pesticides can be efficient, the stratified sampling should provide representation from the full set of strata that have been defined for the pesticide use area.

Conversely, any stratification plan that ignores existing information (i.e., use areas, intensity of application) for the major pesticides or fails to discriminate these pesticides in developing the "vulnerability" of the CWS is highly inefficient from the standpoint of estimating the distribution of concentration for a set of pesticides that should have a priority in the Agency's initial research.

While stratifying by vulnerability is important, random sampling has a feature that allows an inference about the statistical population (a.k.a. "domain" of interest) under the assumption of representativeness. While it may sometimes be inefficient, it is typically reliable. Even though the Panel did raise the advantages of random sampling, the advantage of stratification is that it can focus limited empirical effort on events where regulation can make a difference. Thus, the Panel concluded that stratifying by vulnerability would be much more useful than stratifying by geography.

Therefore, the Panel raised the question - How should vulnerability be defined? A very crude system could be used that calls vulnerable any small or medium CWS in a higher pesticide use intensity region. (A county that is in the top quartile of use for any compound is said to be in a "higher pesticide use intensity area".) It is unclear how this might work in practice. The chance that a watershed is in the top quartile of at least one of 25 chemicals is a Bonferroni problem. Of course it depends on the statistical associations among the target chemicals, but it would be surprising if it is not very close to one. This would mean that virtually all small and medium-sized CWSs are vulnerable.

The Agency's interest appears to be in defining vulnerability in a more refined way. It was claimed that stratifying by vulnerability could give the study design more statistical power to address its goals. But where are the power calculations that suggest this? The Agency has not provided a clear statement of these specific goals. The general goals are clear (although they seem to be growing in number), but the specific inferential or risk-analytical goals are as yet unclear. Stratifying doesn't automatically improve power. In fact, it can reduce it. The Agency would need to articulate the argument for using stratification by vulnerability, or, for that matter, by any criterion.

The Agency's background document asserts that we do not know which variables are likely to be the best indicators of vulnerability. The Agency has suggested that the effectiveness of different criteria for predicting vulnerability will likely vary from chemical to chemical. The Agency's background document suggests that two independent criteria of vulnerability be considered simultaneously. It suggests that this approach would double the chances of getting reasonable and interpretable results and provide twice as much play for the risk managers to design mitigation strategies. Such an approach might have these advantages, but the cost would be to square the sample size required to maintain the desired data quality. It seems unlikely to some SAP members that this would be a workable approach in this context.

It was suggested that a pilot study (or perhaps another pilot study) might be of value to look for a good vulnerability index, as well as to test other design ideas. If scientific knowledge on the subject is as spotty as the Agaency background document suggests, then the best variables and how they should be combined into a vulnerability index is surely an empirical matter, rather than one that can be decided by an expert panel without specific empirical study of the question. However, some SAP members noted that the past extensive process studies and deterministic models developed by USEPA, USDA, USGS, and others can be used and should be consulted before we give up on formulating a vulnerability index. It might still be entirely prudent to test their formulation against the real world in a pilot study. In part, it was noted that it was not clear how past studies may have been used (as pilot studies) to guide the current design. Ensuring adequate use and consideration of extant data is always of benefit.

A pilot study can be simple to design, based on random sampling with as many variables as possible, including geographic variables, chemical-specific variables, transport/fate variables, and other ancillary variables. Exploratory data analysis, including different kinds of discriminant analyses such as traditional linear methods but also non-parametric methods, should allow a much more refined design for the full monitoring effort. Discriminant analysis can specify a vulnerability index and quantify how useful it would be. Of course, using a formal statistical analysis is based on the theory that if it's worth doing, it's worth doing right. A pilot study would likely have many other practical benefits, such as providing an opportunity to work out some technical details and test various reasonable hypotheses that could simplify the entire effort. Now, the result may be that there is no reasonably general definition of vulnerability on which sampling might be usefully stratified. But, of course, it would be very important to know this.

Once the nature of the stratification, if any, has been determined and the specific inferential or risk-analytic goals for the study have been established, the calculations can be carried out to specify the desired number of samples per stratum, or at least to understand the balance between sample size and power. Although this was not explicitly addressed in the presentations, it is presumably the whole reason for stratifying in the first place. It allows optimizing the sampling strategy by focusing on the most vulnerable systems. Notice that shifting samples to vulnerable strata does not mean that we are ignoring the CWSs where the concentrations are not high. Nor would stratification necessarily bias the distribution upward. We still get the entire distribution, with all concentrations high and low. We're just making sure that the estimate of its right tail is especially good.

In some of the Agency's discussion, it was inferred that small CWSs were more vulnerable than larger CWSs. Studies done by SAP members suggested that this was not accurate to generalize to the size of the water system. Smaller watersheds tend to be more vulnerable to higher concentrations of pesticides than larger watersheds, but this does not translate to system size (often measured by the population served or the volume of water produced).

4.) EFED is proposing to use a geographic information systems (GIS) watershed characterization tool for identifying CWSs in the domains and strata.

- **Does the SAP agree with the usefulness of this tool for these purposes?**
- **We are aware of limitations in the accuracy of the data used in the site selection tool. Is the panel aware of better ways of handling the accuracy and precision limitations of this data?**
- **Is the Panel aware of alternative data sources that may not have these limitations?**

The development of a technically sound approach for estimating human exposure to pesticides from surface-derived drinking water depends critically upon the development and use of the proposed GIS watershed characterization tool. The approach appears to be building on extensive national data sets and therefore appears to maximize the use of readily-available information. One SAP member noted that the initial step for this effort should be a characterization of the surface water sources for each CWS using the GIS tool. Thus, an objective and defensible approach to determining how limited resources should be allocated needs to be developed. It was emphasized that this population characterization should be viewed as a first step in the development of a sound monitoring strategy. The FIFRA SAP agrees with the Agency that there are limitations about use of these data.

The availability of new national GIS layers such as the National Hydrologic Data (NHD) and the National Elevational Data (NED) makes this characterization more accurate (and therefore more useful) than ever before. Automatic delineation of watersheds is now computationally feasible, although this may be subject to large error in areas of low relief. A concern is the quality of other types of information and the appropriate scale at which it was collected (e.g. the pesticide use data). This is especially the case as it pertains to small watersheds both to the availability of information at that scale (e.g., 8-digit hydrologic units may not be appropriate in many cases) as well as the accuracy of these data if they are available. There is not an easy resolution to these concerns. The Agency should continue to consult with USGS and USDA for the most detailed and up-to-date sources.

5.) EFED recommends collecting raw water and finished water samples with reactive analysis of the finished water if pesticides are detected in the raw water. Finished water more directly reflects drinking water exposure, while raw water better reflects watershed and usage characteristics allowing better risk management decisions.

- **Given the analytical costs for paired raw and finished water samples and the lack of comprehensive information on water treatment effects on the removal and transformation of most pesticides, does the panel have alternative recommendations on assessing raw and finished water?**

Strong arguments have been made for the value of having both raw and finished water concentration information. Since human exposure is the ultimate goal of this exercise, an appropriate characterization of the human risk requires an understanding of how watershed processes combine to control transport from fields to streams. Thus, both raw and finished water should be tested. This addresses the issue of multiple objectives.

The SAP agrees in principle with the general approach of reactive analysis - to lead with the analysis of raw water and follow up with analysis of finished water. However, some members felt that it was not practical given the large numbers of samples of diverse analytes - some with relatively short sample holding times compared to expected analytical turn-around time. The SAP did express concern about logistical problems related to the functioning of each water-supply system, such as trying to time sampling in the raw and finished water to try to insure that the same parcel of water was being sampled (i.e., by trying to estimate process/holding times through the treatment system). This is necessary to evaluate specific treatment effects. Another idea suggested for consideration was to conduct a unit design study; after the water systems are selected, assess what treatment is in place and then select key, representative systems for paired sampling.

Given the cost constraints, the Agency might consider using some threshold for reactive analysis other than the non-detect level. For example, finished water might only be analyzed if a raw water concentration exceeded 1/10 of the MCL for one of the target compounds. Where to set the threshold would be a difficult balancing act.

6.) Preliminary analyses by EFED indicates pesticide fate properties, sampling frequency, and hydrologic residence time impact the accuracy of calculated annual mean concentrations.

- **Does the SAP have recommendations on balancing the characteristics of the pesticides and the water body being sampled with the number of samples needed to adequately measure the annual mean at different CWSs?**

This is a question that has two dimensions, each with a range of different situations. There is no one-size-fits-all answer to adequately address this problem. Ideally, the final choice of sampling frequency should be determined on a individual basis based on the chemical of interest, the hydrologic system of interest and data quality determination.

The first dimension is that of the chemicals. There are two sets of considerations embedded in this. One is based on pesticide fate properties and this has been discussed in the Agency's background document. The other is the timing and number of applications. For some chemicals on some crops, there is usually only one application. For other chemicals there may be up to six applications a year. For urban areas, there may essentially be a continual application in a given watershed throughout the active growing season. These two factors together (persistence and application) will govern how long of a duration the pesticide source will exist in a given basin. For many agricultural chemicals, particularly in the mid and northern latitudes, the pesticide source will be strongest for a single four-month period (April to July). For other examples, such as pesticides in other environments (urban pesticides and the southern latitudes and the west coast) the pesticide source may be active for two sets of four month periods. Each of these generalized situations ideally might be sampled in a different manner to adequately capture the annual mean concentration.

Given the high cost of analysis, the number of samples that are analyzed should be limited. The chemographs presented in the various documents clearly support the concept of variable sampling intensities for different seasons. *A priori* knowledge on the fate properties (especially half-life, times of chemical application, hydrology, etc.) should be used to optimize sampling intensity.

The second dimension is that of the hydrologic system. Streams and rivers need to be considered separately from reservoirs and lakes. Reservoirs and lakes are probably the easiest situation to sample. The pesticide concentrations will change relatively slowly - as a function of pesticide inputs, pesticide persistence, turnover time, and hydraulic residence times. For long-lived pesticides, the sample frequency may be relatively low and still result in an adequate representation of the annual mean. The ratio of the chemical half-life and hydraulic residence time may provide guidance on the minimum sampling frequency required. Residence time reflects a reservoir-oriented perspective; a much larger problem is short term variability in flowing water systems. Streams and rivers, on the other hand, are much more dynamic than lakes and reservoirs. Rivers and streams form a continuum from drainage ditches to continental rivers. The duration of a specific pesticide peak concentration will vary throughout this continuum. Sampling frequencies will need to be greater in flowing water systems (except the continental rivers), perhaps by a factor of three to four or even more, to obtain comparable confidence intervals around the annual mean concentration.

It might be possible to obtain more information for a given amount of sampling by using daily flow data (at stations where it is available) as an auxilliary variable. This approach is commonly used in pollutant load estimation, either with ratio or regression approaches, and it might be adapted in this context to adjust the mean concentration. Its utility would be primarily in the flowing water systems.

A primary statistical objective of the proposed program is to estimate the national distribution of average annual concentrations of pesticides at the level of the individual CWSs. These data will in turn be used to assess the chronic exposure risks of the populations served by these CWS. The annual sampling program within each selected CWS should be optimized for the estimation of the annual mean concentration for that CWS. Based on the empirical evidence presented, this suggests that the number and temporal distribution of the water samples extracted and analyzed for each site should be tailored to the type of water source, the environmental fate and persistence of the pesticide, and any other observable factors that are known to effect the distribution of the pesticide's concentration over the annual period. The empirical data provide basic guidance on the nature of the annual sampling program for specific pesticides and water source types, e.g., the greater temporal stability in concentrations for reservoirs and lakes as opposed to rivers and streams and the illustrated longer-term persistence of atrazine vs. bromoxynil. The Agency suggests the use of prior simulations of annual concentration distributions for individual pesticides to guide the number and timing of water samples for each selected CWS. This is a very reasonable approach if these simulations are able to reasonably capture the shape of the true concentration distribution; however, care should be taken not to adopt sampling strategies that are not robust against nontrivial departures of the actual from the simulated distributions. This a particular concern for any simulation-based plan that highly concentrates the water sampling within a very narrow time frame of spiking concentration. A disadvantage to developing a sampling plan for water samples from a CWS that is adapted to the simulated distribution of a particular pesticide (e.g. atrazine) is that it may be completely inefficient for other pesticides (e.g. bromoxynil).

Finally, depending on permanent and seasonally-adjusted water treatment practices at the individual CWSs, a sampling plan that is optimized to simulated temporal distributions in concentrations for pesticides in raw water samples may not be optimal for treated, "finished" water outputs.

The Panel was concerned about the inferences in this discussion. One SAP member suggested that in part the Agency was asking about whether (and then how) sample sizes should be differentially allocated when there are different measurement errors associated with data collection. In particular, suppose that measurement error for population A is five times larger than that for population B. How should we apportion the available empirical effort to get estimates of the respective means that have roughly the same precision?

The measurement errors might be different in the two populations for reasons that have nothing to do with any systematic difference between them. For instance, suppose that the laboratory that made the chemical determinations for A had inaccurate protocols and reported error ranges that are 5 times wider than those for B, but not biased either way. Another source of random differences might be that samples from A experience wider temperature fluctuations resulting in more variable degradation histories. If the measurement errors are different because of the random components have different sizes, then there would be an impact on the optimal allocation of sampling effort. Precision can be measured as the reciprocal of the standard error.

7.) Annual mean pesticide concentrations occurring at any CWS vary from year to year. A multi-year study would help to quantify year-to-year variability but is more costly.

- **Does the SAP have any suggestions for assessing annual variability given the financial constraints of the survey?**
- **Would drawing out the survey over three years (with the same number of samples per CWS) improve it?**

It is clear that both spatial and temporal variability are important in characterizing pesticide residues in drinking water. It is clear from the Agency's background document that a significant amount of variability in pesticide concentrations in surface waters will be related to year-to-year changes in the dynamics of climate fluctuations, usage trends, and pest pressure. The answers to the questions asked have more to do with the relative importance (magnitude) of temporal variability than spatial variability.

The SAP is in favor of incorporating some degree of temporal sampling in the study. A multi-year survey effort is required to meet either of the goals of the survey--exposure estimates and model construction. One suggestion from the Agency was to draw the survey out over three years. There are a number of ways of doing this. One approach is to sacrifice spatial coverage by reducing the number of CWSs surveyed to a third of that proposed, with each CWS sampled for a longer period of time. Another approach is to simply sample only one-third of the total CWS in any one year with a non-overlapping set in any one year, though this approach could loose information on year-to-year changes for individual systems.

With no apparent historical data on year-to-year trends in pesticide levels for a wide range of CWSs, there appears to be little that can be done except to extend the study for several years to collect data to estimate year-to-year variability in annual average pesticide levels and upper percentiles of pesticide concentration distributions in CWSs. The number of extra years to collect data will be a function of the cost and the amount of year-to-year variability that is observed in CWS pesticide levels. The issue of whether the particular year of a one year study is representative or not could be evaluated by comparison with longer-term data series such as NAWQA or the Heidelberg College data from Ohio. Long-term precipitation and flow information would also offer a useful perspective, though this approach could use information in year-to-year changes for individual systems.

One issue that should be addressed is to not only monitor CWSs in future years that exhibited high pesticide concentrations but a random sample across all CWSs in the baseline study since CWSs with high concentrations may be either in high years or just be at elevated levels in comparison to other CWSs in the baseline study. The subset of CWSs to continue monitoring in future years should also include CWSs that had low and moderate levels of pesticides in the baseline study as they may just be in a low range of a temporal trend and future years will exhibit higher pesticide levels.

Instead, some form of interpenetrating survey design might be used. In a simple example, the total number of CWSs is reduced by a fraction, say 20%, and an equal fraction of sites are selected for multi-year sampling, keeping the total number of samples the same. If a longer time view is selected, that is a true monitoring plan is desired, CWSs could be sampled for a couple of years then replaced with other CWSs with no one CWS being in the sample for more than say 3 years. This allows the monitoring plan to cover the whole region fairly uniformly over time while also collecting the needed information on temporal variability.

The question to the Agency is this, "How do you plan to interpret the resulting distributions of parameter estimates if no temporal sampling is performed?" In agricultural research, it is traditional that studies be replicated over three years, five is preferred, if the researcher wishes to make believable inferences to expected results over time. Something similar to this needs to be incorporated into this study as well. Uncertainty estimates in annual average pesticide levels in CWSs would be underestimated and lead to biased confidence intervals for parameters of concern in this study if year-to-year variability is not included.

Even if additional funding is not provided, the Agency should not adopt a single year sampling scheme. The resulting exposure estimates will not be credible because of the significant year-to-year variability in the factors that contribute to pesticide occurrence. Three years should be the minimum duration considered; five years would be preferred.

8.) EFED recognizes that pesticide concentrations in drinking water are dependent on factors including watershed characteristics, pesticide use, pesticide fate properties, surface water hydrology, and water treatment processes. Interpretation of the monitoring data will be dependent on the collection of such related ancillary data.

- **What types of ancillary information does the SAP believe would assist in the interpretation of the monitoring data, and application of the data to model development and validation?**

While exposure estimates could be made without use of ancillary data, model construction requires it. The Agency has identified a reasonable set of watershed properties and ancillary data. A greater concern expressed by the SAP was not which data but the quality and scale of these data. It is critical that these be collected as part of the study. Development of models (and their successive refinements) is an important component for determining vulnerability. An aspect that hasn't been discussed is to validate the data layers used in developing models--in other words, how accurate are the data derived from county-level crop surveys when aggregated by watershed. Much of the relevant information (e.g., physical characteristics of the watershed) may already be available in some electronic format, but it may need to be verified and adjusted for scale. Some desired information is perhaps not readily available and expensive to obtain during the years of the study, such as: actual pesticide usage, actual timing of application, agronomic measures, crop types, weather (esp. daily precip, temp, wind), and hydrologic (stream discharge) data. Perhaps this detailed data collection should be for selected subsets of CWSs, and the more general the ancillary data collected for all sites.

It is important that historical records on pesticide monitoring at each selected CWS be obtained and used to augment the sampled data. Estimates of pesticide use for each watershed will also need to be obtained. USDA should be consulted on how they may be able to help provide use data. These data collection costs need to be estimated and included in the program budget prior to implementation of the monitoring program. A significant portion of the funding for the project may need to be allocated up front to cover the costs of obtaining and analyzing this information.

9.) We have defined our population as CWSs with pesticide use in their watersheds, based on the assumption that runoff and near field spray drift are the major routes of loading to the water supply.

- **Is this a reasonable assumption and should we monitor facilities that do not have pesticide use in the watershed?**

In general, defining the population of pertinent CWS by pesticides used in their watersheds is appropriate. It should be realized that there are a number of situations where pesticides are used or introduced into the watershed beyond that which is reported in agricultural use statistics --the data that would be used to estimate pesticide use for the watersheds. These situations include illegal use, non-agricultural use (e.g. urban use, forestry, aquaculture, roadways), manufacturing/distribution, and atmospheric deposition. Chemical loading from groundwater discharge to streams is also a component, but this would be most significant from within the watershed.

In the FIFRA SAPs experience, and from their review of the scientific literature on pesticide use in forestry and aquaculture, long-range atmospheric deposition generally can be considered insignificant for the purpose of this survey. While pesticide release during manufacturing/distribution could likely be considered insignificant for the purpose of this survey, as well, the locations of pesticide manufacturing facilities should certainly be understood in relation to sampling locations. They might be purposefully excluded or included but their potential impact shouldn't be allowed to confound the results. Illegal use is always an unknown. Non-agricultural use statistics generally do not exist, as is reported in the Agenecy background document. Roadway use could be important for some watersheds, but the use varies greatly from state to state. Urban use of pesticides may have a substantial effect on their stream concentrations. As expected, USGS studies suggest the environmental fate of agricultural and urban use pesticides differ.

The Agency's background document correctly calls for inclusion of CWS that could be influenced by urban areas to be included in the survey. Finally, prometon may be in a unique situation. It is widely used for non-agricultural purposes in both agricultural and urban environments. It has been widely observed in streams across the United States, but is not listed on any of the standard pesticide use surveys. All of these examples point to the limitations of using agricultural pesticide use alone as a determinant of the CWS populations, but also suggest that the resulting errors will not be too great. Of course, as part of the basic design a few reference (background) watersheds, with low or no pesticide use, could be included as a quality assurance measure.

10.) One option proposed in the design framework is a census of the facilities serving the largest cities. These facilities are believed to have little agriculture in their watersheds, but reflect the drinking water for a large percentage of the population.

- **Does the SAP believe sampling CWSs serving very large populations is useful even if they are expected to be of low vulnerability?**

The SAP agrees that sampling the large population CWSs is necessary. One of the main objectives of the study is to determine the human exposure to pesticides from surface waters. Therefore, sampling CWSs serving the very large populations should be carried out with an inclusion probability of 1.0. If sampling of large population CWSs is not implemented, it would be an assumption in this study that these CWSs have low levels of pesticides and impose an insignificant human exposure.

The question of whether to sample or not sample large population CWSs is an issue of science policy and must balance the risks associated with making decisions based on validated or un-validated assumptions. However, for CWSs where there is verification of little use and occurrence, sampling might be performed at a reduced frequency. This would also seem dependent upon the choice of primary sampling domain. If pesticide use is the primary domain, and none of the largest systems are in the domain for that pesticide, then they may not be necessary. Also, the large systems often do more complete sampling and monitoring. There may already be data for at least some compounds. It could at least guide a choice of frequency. Large systems might be willing to bear the burden of some limited sample collection and monitoring. Some balanced approach will be necessary, because large systems are complex and the monitoring could be very costly if not designed carefully.

11.) Modeling should allow the Agency to reduce the extent of future drinking water surveys, and help to better identify areas at higher risk.

- **Are there any changes to the design which will better allow the survey to support model development and testing?**

The Agency's proposal to develop a stratified random sample of CWSs with oversampling of

CWSs expected to have higher concentration levels ("vulnerability") is a good design to also begin acquiring the empirical data needed to begin building and testing models to predict pesticide concentrations in raw and finished drinking water. At this stage in the research process, some SAP members feel we have neither a well-developed theoretical understanding nor the ancillary data needed to specify good models for single pesticides. The proposed survey in combination with existing and new ancillary data for sampled CWSs and their source watersheds should provide a good data set with which to begin preliminary model development and testing. The immediate goal should be to develop and test preliminary models for a high-priority set of pesticides. The Agency should also be thinking ahead to begin building a framework for how measured compounds will be used to proxy model behavior for unmeasured compounds. This planning may well guide the choice of which representative pesticide compounds from general classes should be the focus of initial model development and testing. Finally, to build initial models of the detailed multivariate and spatial form described by the Agency will require new ancillary data for model projections. New survey data or cross-validation based on the existing samples (i.e. building models with half the CWSs, testing the performance on a second half) will be needed to evaluate the performance of the initial modeling effort. Ultimately, if acceptable models can be defined, broader collection of ancillary data will be needed if the model is to be applied generally to predicting chronic exposures to pesticides over time and in other CWSs.

Related to improving models, better models are made by challenging the model predictions. This happens when the situation being considered represents an extrapolation beyond the range of conditions used to build the model. If the survey can extend the range of conditions beyond those currently predicted well by the existing models, then the models can be challenged and when necessary improved. Little new is learned when data are collected where we know the model works well.

If supporting model development and testing is a priority goal, stratification based on model predictions may also be warranted. As pointed out in the International Life Sciences Institute (ILSI) model report, the goals and data needs of model calibration and validation do not always coincide with those of developing good concentration distributions. This suggests there may have to be either a trade-off in design efficiency to accommodate both objectives, or one objective may need to be dropped, dependent on funding.

However, some of the SAP members don't agree with the ILSI findings that regression-based models have utility only in the lowest tier of assessments and that high precision can only come with detailed process models. Empirical modeling, of which regression is one approach, has resolution that is dependent on the amount of data available: the more data available, the higher the resolution. In this case, a lot of data will be generated, and there is the potential for empirical models to be quite effective for higher tier assessments without going to the more complex process models.

The SAP was concerned that at the proposed level of funding, it may not be possible to collect the ancillary data needed for model development to meet the reliability criteria needed for regulatory purposes AND estimate exposure for 10-20 pesticides. Yet, without the model development, ongoing monitoring will be required to estimate exposure in future years and estimate exposure for pesticides excluded from the priority set. The Agency must carefully evaluate these objectives and the trade-offs in planning and evaluating the funding necessary to collect the information needed to produce reliable models. Process-based models combined with GIS tools may also have a useful role, particularly in understanding areas at higher risk and predicting the behavior of new compounds.