SAP Minutes No. 2004-04

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

A MODEL COMPARISON: DIETARY AND AGGREGATE EXPOSURE IN CALENDEX, CARES, AND LIFELINE

APRIL 29 and 30, 2004
FIFRA Scientific Advisory Panel Meeting, held at the Holiday Inn Rosslyn at Key Bridge, Arlington, Virginia

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 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 Myrta R. Christian, SAP Designated Federal Official, via e-mail at christian.myrta@.epa.gov.

In preparing the meeting minutes, the Panel carefully considered all information provided and presented by the Agency presenters, as well as information presented by public commenters. This document addresses the information provided and presented by the Agency within the structure of the charge.

CONTENTS

SAP Minutes No. 2004-04

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

A MODEL COMPARISON: DIETARY AND AGGREGATE EXPOSURE IN CALENDEX, CARES, AND LIFELINE

APRIL 29 and 30, 2004
FIFRA Scientific Advisory Panel Meeting, held at the Holiday Inn Rosslyn at Key Bridge, Arlington, Virginia

Myrta R. Christian, M.S.
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: July 29, 2004

Stephen M. Roberts, Ph.D.
FIFRA SAP, Session Chair
FIFRA Scientific Advisory Panel
Date: July 29, 2004

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
April 29 and 30, 2004

A MODEL COMPARISON: DIETARY AND AGGREGATE EXPOSURE IN CALENDEX, CARES, AND LIFELINE

PARTICIPANTS

FIFRA SAP Chair

Stephen M. Roberts, Ph.D., Professor & Program Director, University of Florida, Center for Environmental & Human Toxicology, Gainesville, FL

Designated Federal Official

Myrta R. Christian, M.S., FIFRA Scientific Advisory Panel Staff, Office of Science Coordination and Policy, EPA

FIFRA Scientific Advisory Panel Members

Stuart Handwerger, M.D., Director, Division of Endocrinology, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH

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

FQPA Science Review Board Members

Bernard Engel, Ph.D., Professor, Agricultural & Biological Engineering, Purdue University, West Lafayette, IN

Scott Ferson, Ph.D., Senior Scientist, Applied Biomathematics, Setauket, NY

Natalie Freeman, Ph.D., Adjunct Associate Professor, Environmental/Community Medicine, Robert Wood Johnson School of Medicine, Piscataway, NJ

Chensheng Lu, Ph.D., Department of Environmental and Occupational Health Sciences, School of Public Health and Community Medicine, University of Washington, Seattle, WA

Peter D.M. Macdonald, D.Phil., Professor of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada

David MacIntosh, Ph.D., Senior Associate, Environmental Health and Engineering, Inc., Newton, MA

Kenneth Portier, Ph.D., Associate Professor, Statistics, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL

Nu-may Ruby Reed, Ph.D., D.A.B.T., California Environmental Protection Agency, Department of Pesticide Regulation, Sacramento, CA

P. Barry Ryan, Ph.D., Professor, Environmental & Occupational Health, Rollins School of Public Health, Emory University, Atlanta, GA

Michael D. Sohn, Ph.D., Scientist, Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, CA

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 model comparison: dietary and aggregate exposure in Calendex, CARES, and LifeLine.

Advance notice of the meeting was published in the *Federal Register* on March 24, 2004. The review was conducted in an open Panel meeting held in Arlington, Virginia, on April 29 and 30, 2004. Dr. Stephen M. Roberts chaired the meeting. Mrs. Myrta R. Christian served as the Designated Federal Official.

A major requirement of the Food Quality Protection Act is that exposures to pesticides across various pathways and routes (e.g., oral exposure through food and water, dermal exposure through turf uses, etc.) be appropriately combined such that an "aggregate" exposure assessment can be performed. Over the last several years, OPP has sponsored several presentations to the FIFRA Scientific Advisory Panel of three major pesticide exposure models which have this aggregation capability: DEEM/Calendex, CARES, and LifeLine. These models all permit a time-based integration of both residential and dietary (food and water) exposures to pesticides. This is performed probabilistically such that aggregation (or combining) of residues across multiple routes is accounted for in an appropriate and realistic manner.

Past Scientific Advisory Panel sessions have involved presentations of these three models on separate occasions and resulted in valuable feedback and suggestions. The models have been significantly upgraded and improved, in part, as a result of these SAP comments. The purpose of this two-day session is to present the three models jointly and compare exposure estimates and results generated by each of the three models using a common dataset for a hypothetical chemical. Pathways which will be considered are ingestion through food and ingestion through water.

In contrast to previous SAPs where the model developers were invited to make their presentations to the SAP, this SAP presentation will demonstrate the results of EPA's own model runs. A comparison/contrast regarding the methods by which the models consider and use the data will be considered. Resulting model outputs will be presented and discussed. The main focus of the presentation will be comparison of dietary exposures (food + water) as estimated by each of the three models. OPP plans further comparisons of exposure through all three exposure routes considered jointly (food, water, and residential lawn use) at some later time in the future.

In preparing these meeting minutes, the Panel carefully considered all information provided and presented by the Agency presenters, as well as information presented by public commenters. This document addresses the information provided and presented at the meeting, especially the response to the charge by the Agency.

PUBLIC COMMENTERS

Oral statements were presented as follows:

Stephen Petersen, on behalf of Durango Software, LLC

Christine F. Chaisson, Ph.D. on behalf of The LifeLine Group

1. **General Approach of Approximating Models**:

A. While the three probabilistic risk assessment models described in this SAP presentation each project pesticide exposure for the 'US population', they differ in their basic design in a number of ways. EPA has identified and investigated four model design features associated with these models, as follows:

- 'Reference Population':
  - DEEM-Calendex is based on the CSFII survey design.
  - CARES is based on the US Census PUMS (Public-Use Microdata Samples).
  - LifeLine is based on the NCHS Natality database.

- 'Binning food diaries' to generate longitudinal consumption profiles:
  - DEEM-Calendex draws from the individuals' two day diaries.
  - CARES uses the Gower dissimilarity index.
  - LifeLine 'bins' food diaries based on age and season.

- 'Model weights' to project simulated exposure days up to the modeled (US) population
  - DEEM-Calendex uses the CSFII survey weights.
  - CARES uses weights developed from its stratified sampling design.
  - LifeLine weights each person equally.

- 'Body weight'
  - DEEM-Calendex assigns food consumption values to each individual on the basis of the grams food/kg body weight as reported by the CSFII respondents.
  - CARES also assigns food consumption values to each individual on the basis of the grams food/kg body weight as reported by the CSFII respondents. However, since the CARES body weights are different from the CSFII Body weights, CARES adjusts the amount of food consumed to reflect the CARES body weight.
  - LifeLine uses a reported consumption value for each individual on the basis of the grams food as reported in CSFII) and estimates the individual's body weight based on physiometric growth models developed for various demographic groups (based on gender, race and ethnicity) using NHANES data.

Question 1.1 **The SAP is asked to please comment on whether the above cited model design features reflect those most likely to result in**

differences in dietary [food and water] exposure estimates based on identical data sets. If not, what other model design features are likely to cause different dietary exposure estimates?

B. In an attempt to further elucidate differences between predicted exposures among the three models (DEEM-Calendex, CARES and LifeLine), OPP developed SAS approximation models. These SAS approximation models permit the isolation of factors related to the Reference Population, Binning Procedures, Sampling Weights, and individual Body Weights which cannot be isolated by running the individual models. Section IV of the background document, provided to the SAP, describes the development of these SAS approximation models and some analyses performed by the Agency using these SAS approximation models to compare and contrast model design features of DEEM-Calendex, CARES, and LifeLine. Based on these analyses the Agency concluded that the SAS approximation models track actual model results very closely for single Raw Agricultural Commodity (RAC) analyses, and reasonably well for the multi-RAC analyses.

Question 1.2 **The SAP is asked to please comment on the approach taken by the Agency to develop and use SAS approximation models (see Section IV of the background document) to attribute differences in model predictions from differences in model designs. Please suggest possible improvements or refinements to these SAS approximation models and to alternative methods for comparing model predictions.**

2. <u>Reference Population & Model Weights</u>:

The DEEM-Calendex program uses the CSFII survey respondents as its reference population; as such, the DEEM-Calendex model estimates use the CSFII-specific sample (or model) weights to estimate exposures. Each simulated day is weighted to project that exposure day to represent a group of similar individuals from the U.S. population. CARES and LifeLine use alternative data sources (i.e., U.S. Census PUMS, and NCHS Natality) to generate their respective Reference populations. The CARES model developed its Reference Population by taking a stratified random sample of 100,000 persons from the US Census PUMS. The stratified sampling design enabled CARES to over-represent sub-populations of interest (e.g., 20,003 Infants) in its reference population which are subsequently downweighted to permit projection to the U.S. population. The LifeLine model uses the Natality data to generate its Reference population. LifeLine provides the option of using CSFII survey weights to affect the probability of selecting diaries from each of the dietary bins. If this option is not selected, LifeLine will weight each modeled individual equally since these modeled lives are drawn randomly from the Natality statistics.

Question 2.1 **The SAP is asked to please comment on the different approaches used by the three models in developing their Reference Populations and model weights.**

3. Binning Design & Frequencies of Using CSFII Diaries:

These models differ in the expected (or actual) frequencies that each CSFII diary is used in the probabilistic risk assessment. DEEM-Calendex uses only the individuals that provided two days of food diaries in its reference population, and sets aside approximately 1,000 one-day food diaries in estimating dietary exposure. CARES employs a Gower dissimilarity index in its algorithm to generate longitudinal consumption profiles for its Reference Population. The result is use of some CSFII diaries much more often than other diaries in simulating exposure (as much as 4,000 times for certain diaries versus once for others). Approximately 1,000 CSFII diaries are not included in the CARES Food Match table. The LifeLine model uses a very general bin based on age and season, such that all food diaries within a particular bin have the same expected frequency of being used in its exposure assessment. In order to evaluate the effect of these differing frequencies and modeling weights, EPA approximated all three models using the LifeLine recipes (i.e., keeping recipes constant).

Question 3.1 **The SAP is asked to please comment on the frequency that CSFII diaries are used by the various models. Are there any potential biases that may arise in the respective dietary exposure estimates for these models as a result of how they used CSFII records? Considering LifeLine's current dietary bin design (age, season), please comment with respect to the use of the CSFII survey weight option. Is either LifeLine option (CSFII-weighted or not) generally more appropriate than the other or are there circumstances in which one might be preferable to the other?**

4. Commodity Exposure Contribution Analyses:

An important aspect of any dietary risk assessment is the ability to identify significant contributors at the upper percentiles of exposure. The CARES and DEEM models both include an output report option known as the Critical Exposure Contribution (CEC) analysis. A comparable report option is expected to be developed for the LifeLine model in the near future. These CEC reports quantify the contribution of specific food commodities (RAC-FF) to the total exposure at the upper percentiles (e.g., top 0.2%) of the exposure distribution. An alternate or complementary approach (frequency-exceeded), also used by various model developers, tabulates the frequency that a particular commodity (RAC-FF) causes exposure to exceed some level of concern. As was the case with predictive exposure estimates, model design can affect the outcome of commodity exposure contribution analyses. Section IV.G of the background document

describes the CEC and 'frequency-exceeded' approaches for identifying significant contributors at the upper end of the exposure distribution. Tables 13 and 14 show CEC reports and 'frequency of occurrence' data for DEEM-FCID and CARES analyses for 3 - 5 year olds and 20 - 49 year olds, respectively. Tables 15 and 16 show SAS approximations for the model CEC reports and 'number of occurrences > aPAD' for these same age groups. Although there is certainly a degree of similarity between model results and between the model results and the SAS approximation results, differences do occur.

Question 4.1 **The SAP is asked to please comment on the relative merits of the two approaches described above (CEC and frequency-exceeded) for identifying significant contributors (RAC-FF) to exposure at the upper percentiles of exposure. Are there other methods or techniques which the Panel might recommend for accomplishing this important part of the dietary exposure assessment?**

SUMMARY OF PANEL DISCUSSION AND RECOMMENDATIONS

The FIFRA SAP reviewed the Agency background documents relating to a comparison of three models; Calendex, CARES and LifeLine. All have been proposed for use in dietary and aggregate exposure assessments and have been examined separately in past FIFRA SAP sessions. The purpose of this review was to comment on the methodology and results of an analysis performed by EPA OPP specifically to compare estimates generated by each of the three models using a common dataset of a hypothetical chemical through food and water ingestion pathways.

The EPA OPP analysis focused on several important aspects of the three models, namely differences in the reference population used, how food diaries were binned, use of sampling weights in simulating exposure days and how body weights were incorporated into the exposure calculations. The analysis demonstrated similarities and differences. A number of limitations were noted by the SAP, namely that the analysis was limited to a one-day exposure scenario and that factors that are known to impact dietary exposure besides age, such as season and geographic location are not considered. The SAP acknowledged that the current lack of data on longitudinal food consumption is a major impediment to implementation of multi-day assessments.

The SAS models used appeared to be adequate for the analysis objectives but additional documentation is needed. The SAP found it difficult to compare model predictions because the associated uncertainty was not provided. The SAP recommended use of Bayesian and Monte Carlo methods for this task. It was suggested that better comparisons of the models might be accomplished using the median or 75^{th} percentile instead of the 99^{th} percentile concentrations.

What was striking about the results presented is that although there are differences in the construction of the underlying reference populations and each model's use of sampling weights, the differences in predicted concentration at the 99th percentile appear to be minor. Since all the models are in essence accounting programs, the small differences observed would not be the consequence of differences in underlying mathematics. Limitations and potential biases in the surveys from which the reference populations are drawn were discussed as one reason for this lack of differences. Another could be the strong dependence of all models on the CSFII data. Biases in the use of the CSFII data by the different models were also discussed, including the uncertainty imposed by some CSFII diaries not being used by some/all models. The SAP recommended that future studies concentrate on testing each model's specific reference population design. Alternate methods of binning diets were discussed, such as using sampling without replacement and incorporating seasonality. In addition, the SAP recommended that identification of the high-end exposed individuals/diets be added to the critical exposure commodities and frequency-exceeded analyses currently provided by the models.

The SAP recommended that EPA OPP continue to use all three models as one method of incorporating model uncertainty into an assessment. The SAP considers it valuable to continue the process of understanding the strengths and weaknesses of each model, and to possibly include simple statistical and mechanistic models in the comparisons as well. The development of the SAS models used in this analysis was seen as a good start.

PANEL DELIBERATIONS AND RESPONSE TO CHARGE

The specific issues addressed by the Panel are keyed to the Agency's background documents and the Agency's charge questions.

Response to Charge

1. <u>General Approach of Approximating Models</u>:

A. While the three probabilistic risk assessment models described in this SAP presentation each project pesticide exposure for the 'US population', they differ in their basic design in a number of ways. EPA has identified and investigated four model design features associated with these models, as follows:

- 'Reference Population':
  - DEEM-Calendex is based on the CSFII survey design.
  - CARES is based on the US Census PUMS.
  - LifeLine is based on the NCHS Natality database.

- 'Binning food diaries' to generate longitudinal consumption profiles:
  - DEEM-Calendex draws from the individuals' two day diaries.

- CARES uses the Gower dissimilarity index.
- LifeLine 'bins' food diaries based on age and season.

- 'Model weights' to project simulated exposure days up to the modeled (US) population
  - DEEM-Calendex uses the CSFII survey weights.
  - CARES uses weights developed from its stratified sampling design.
  - LifeLine weights each person equally.

- 'Body weight'
  - DEEM-Calendex assigns food consumption values to each individual on the basis of the grams food/kg body weight as reported by the CSFII respondents.
  - CARES also assigns food consumption values to each individual on the basis of the grams food/kg body weight as reported by the CSFII respondents. However, since the CARES body weights are different from the CSFII Body weights, CARES adjusts the amount of food consumed to reflect the CARES body weight.
  - LifeLine uses a reported consumption value for each individual on the basis of the grams food as reported in CSFII and estimates the individual's body weight based on physiometric growth models developed for various demographic groups (based on gender, race and ethnicity) using NHANES data.

Question 1.1 **The SAP is asked to please comment on whether the above cited model design features reflect those most likely to result in differences in dietary [food and water] exposure estimates based on identical data sets. If not, what other model design features are likely to cause different dietary exposure estimates?**

Response

The Agency staff demonstrated a detailed understanding of the DEEM-Calendex, CARES and LifeLine models. The features that are most likely to result in differences in dietary exposure estimates from the three models were identified based on the work presented in the documents and in the presentations. The results seem to indicate that the models work as intended. The results presented also suggest that the models provide similar estimates at the Agency-determined levels of interest. Each of these models should continue to be used since they possess unique features that will prove useful in looking at different issues and more complex questions. The Agency staff should also continue to identify and develop other models as new features or needs arise.

Of the design features listed, the most important is the difference between generating lives from birth and sampling from a given population. It was suggested that a comparison of the demographics of the reference populations generated by the different models be included. For instance, the Natality data set used to populate the LifeLine model is a cohort of infants born in 1996. This cohort is likely to be more representative of the child-bearing population in the US at that time rather than the US population as a whole (and represented by the US census).

The model features presented appear to be reasonable approaches to generating estimates. However, there is little information available to assess the validity of the analyses. There is little "ground-truthing" in the assessments. Efforts to identify data, even at intermediate stages in an analysis, for model comparison as well as validation will be desirable to further demonstrate the validity of the models. Additional details on potential ground-truthing efforts are provided in the response to question 4.

The SAP recognizes the challenges that are faced in moving from a one-day to a multiple-day analysis and offers the following suggestions.

It is important to note that the Agency presented to the SAP a comparison of one-day food consumption distributions generated by the three models as opposed to modeled distributions of dietary pesticide exposure. Differences in one-day food consumption records as simulated by the models are anticipated to be small compared to the differences likely to result from alternative approaches to estimating time-integrated or multi-day exposures, as well as from variability in pesticide residue concentrations. In addition, bodyweight is expected to have minimal effect on the population variance of pesticide dietary intake. Nevertheless, the apparent strict relationships between the food portion size and body weight used by some of the models should perhaps be relaxed to better reflect real-world dietary practices.

Ideally, the models should incorporate factors that reflect differences in dietary exposure estimates. Besides age, which is used in all three models, factors such as season and geographic location where the individuals reside should also be considered in the models. Considering the relatively limited food diaries in the CSFII, incorporating additional factors will reduce the statistical power of the model's estimating ability because of sample size limitations. The impact of reduced sample size must be considered when deciding to add a particular factor to a model.

The models appear to behave similarly at the extreme tails, possibly because they are based on the same unique cases. In addition, model misspecification errors could also drive the results at these tails. Using the full uncertainty distributions for these upper percentiles when comparing model predictions would help determine which of these two is actually driving the extreme events.

Under the constraints of this question, the SAP was forced to differentiate between model features by analyzing predictions at the very tail of an uncertainty distribution. The

assessments shown by the EPA, under this criterion, appear to show similar results. However, the tails of an uncertainty distribution may not be an appropriate criterion for comparing the three models, especially when examining multiple day exposures. A fuller suite of summary statistics should be examined.

A number of challenges were identified associated with moving from single day exposure estimates to multiple day estimates with existing data and approaches. The lack of longitudinal data on food consumption remains one of the more serious data deficiencies. In addition, the models do not account for likely correlations of pesticide residue concentrations associated with the same food consumed on subsequent days by the same individual. The current approach in such cases is to draw a new pesticide residue concentration for each day. It was noted that this same problem (same food consumed multiple times having pesticide concentrations that are correlated because the foods were purchased from the same source) exists within the current one-day exposures presented.

B. In an attempt to further elucidate differences between predicted exposures among the three models (DEEM-Calendex, CARES and LifeLine), OPP developed SAS approximation models. These SAS approximation models permit the isolation of factors related to the Reference Population, Binning Procedures, Sampling Weights, and individual Body Weights which cannot be isolated by running the individual models. Section IV of the background document, provided to the SAP, describes the development of these SAS approximation models and some analyses performed by the Agency using these SAS approximation models to compare and contrast model design features of DEEM-Calendex, CARES, and LifeLine. Based on these analyses the Agency concluded that the SAS approximation models track actual model results very closely for single Raw Agricultural Commodity (RAC) analyses, and reasonably well for the multi-RAC analyses.

Question 1.2 **The SAP is asked to please comment on the approach taken by the Agency to develop and use SAS approximation models (see Section IV of the background document) to attribute differences in model predictions from differences in model designs. Please suggest possible improvements or refinements to these SAS approximation models and to alternative methods for comparing model predictions.**

Response

The SAS models were considered as reasonable methods for checking the consistency of the DEEM-Calendex, CARES and LifeLine models. A number of reasons for taking this approach were apparent. Earlier SAPs had suggested the approach of coding approximations of each model on a common independent platform. These SAPs had issues regarding the need for some of the complexity included in the models and felt that an independent platform assessment would help. A second objective for developing the

SAS model was to explore comparisons among the models regarding features that were not able to be studied directly in LifeLine.

The flexibility and transparency of the SAS models are important aspects of the assessment of this approach. However, the limited information/documentation of the SAS models provided in the report to the SAP and the Agency presentation made it difficult to determine the degree of approximation incorporated into the code. Future reports should include additional documentation of the SAS model and code. One Panel member characterized the approximation as created by mixture distributions based on observed consumption rates weighted by the number of people in the total population represented by the sampled person. If this is the case, it was not terribly surprising that the approximations matched what the three models produced (since they were designed to estimate the same quantities but by a different approach). In the future, it will be very important to keep the models transparent in both the approach to its development and in its application.

It is natural when asked to compare a table of numbers that purport to describe superimposed distribution functions to first ascertain the uncertainty associated with the numbers presented. The calculations displayed in the comparisons with the SAS approximations were presented as exact, having no associated uncertainty. This was not altogether true though, as they merely *expressed* no uncertainty. There is a larger context of uncertainty that comes from the measurement error and sampling uncertainty associated with the original data on which these calculations were based. The Panel felt it was asked to comment on the importance of observed model outcome differences without being supplied the (uncertainty) context in which to consider these differences. Future SAS model analyses can provide this context and, as a result, will be very useful in identifying inherited biases and uncertainties in the more complex models. The Panel suggested 1) calculation and examination of descriptive statistics at the lower percentiles, such as the 75^{th}, the median, and the 25^{th}, 2) examination of the validity and importance/cause of the spread in outcomes observed at the 90th percentile that is bracketed above and below this percentile by very similar model outcomes 3) extension of the SAS code to incorporate a 2-D Monte Carlo analysis and 4) consider utilizing Bayesian techniques (Lin et al., 1999 and Gelman et al., 1995) to capture the variability in the two-day diet samples. Since the SAS model is an approximation of the other models, it is not clear at which percentile differences between the SAS model and the other models should appear. The suggestions above should help in this regard.

The success of the SAS approximations begs the question as to whether these simpler models themselves might be good enough for some future risk assessments. It is conceivable that the SAS models might be extended to cover cases that the major models handle differently or cases that not all models can handle. It was felt that in general the SAS models should not be used as a replacement for the three more complex models until further review of the SAS approximations.

2. Reference Population & Model Weights:

The DEEM-Calendex program uses the CSFII survey respondents as its reference population; as such, the DEEM-Calendex model estimates use the CSFII-specific sample (or model) weights to estimate exposures. Each simulated day is weighted to project that exposure day to represent a group of similar individuals from the U.S. population. CARES and LifeLine use alternative data sources (i.e., U.S. Census PUMS, and NCHS Natality) to generate their respective Reference populations. The CARES model developed its Reference Population by taking a stratified random sample of 100,000 persons from the US Census PUMS. The stratified sampling design enabled CARES to over-represent sub-populations of interest (e.g., 20,003 Infants) in its reference population which are subsequently downweighted to permit projection to the U.S. population. The LifeLine model uses the Natality data to generate its Reference population. LifeLine provides the option of using CSFII survey weights to affect the probability of selecting diaries from each of the dietary bins. If this option is not selected, LifeLine will weight each modeled individual equally since these modeled lives are drawn randomly from the Natality statistics.

Question 2.1 **The SAP is asked to please comment on the different approaches used by the three models in developing their Reference Populations and model weights.**

Response

When these three approaches were evaluated from a statistical point of view, the SAP came to the following observations/conclusions.

The DEEM-Calendex model simulates its reference population by direct use of the CSFII dataset and hence ties the reference population directly to the diet data base. This reference population is well defined, and has carefully crafted sample weights allowing the model to extrapolate exposure to the full population. All the difficult statistics (computing the sampling weights) have been done outside the model.

The CARES model uses the US Census PUMS (5 million) database as the foundation, but creates its own reference population via sampling from this data base. The CARES Population Generator provides the user a standard 100,000 member reference population. Each observation has an appropriate sampling weight that is used in the extrapolation of exposure to the full population. The fact that the Population Generator is outside the main CARES application implies that the analyst can create alternative stratified sampling protocols and hence has the flexibility to create different 100,000 member reference populations and associated sampling weights. This has the potential for the analyst to create and examine reference populations that have characteristics that are potentially different from the whole US population. For example, specific geographic sub populations could be examined without reducing the total number of individuals in

the reference population (as might be the case in subsetting the CSFII data set). This might be useful if chemical use were regionally specific. Like the DEEM-Calendex model, the census data are cross-sectional views of the population.

The LifeLine model defines its reference population via a set of parameterized models and envisions this reference population as a birth cohort rather than as a cross-sectional snapshot. In their model, the reference population is envisioned as a set of equations indexed by unknown parameters. These unknown parameters are estimated using best available data, and uncertainty in the parameter estimates can in many cases be defined. These parameter estimates are used in the equation in a Monte Carlo simulation to generate the reference population. Since the uncertainty in the model is focused on the estimated parameters, sensitivity analyses are somewhat easier to accomplish. On the other hand, appropriateness of this approach relies on the validity of assumptions of stationary age/sex/race distributions and stability of birth and mortality rates. The model can be modified to handle specially defined sub populations.

With this said, it seems that the DEEM-Calendex model is probably the most representative of the population at the time the data was collected. CARES has the potential to have a reference population that is most representative of the population at the latest census. LifeLine has a reference population that is most idealized and hence is farthest from an actual population. But, because of this, the LifeLine model has the flexibility to examine exposure under different future scenarios.

It was noted that all three reference populations are needed to address different issues. Studies based on the Natality data base are prospective whereas PUMS and CSFII data bases produce cross-sectional studies. The Natality data base will tend to assign higher relative sampling weights to short-lived people, while the PUMS and CSFII will tend to assign higher relative sampling weights to long-lived people.

What is striking about the results presented is that although there are differences in the reference populations and model weights used in the three models, the differences in the results appear to be very minor. It was noted that a possible reason for the lack of differences in the results is that the biases in reference populations are similar in terms of who are included and who are excluded. Individuals who tend to be excluded include émigrés, Blacks, and Hispanics. These similar biases may create similar results at the upper ends of the distribution.

To give examples, CARES uses census PUMS which were drawn from a one fifth subset of the US Census that completed the US Census long form. When weighted by within state populations, the PUMS are reasonably representative of the whole population. There is at least one article (Ong and Ong, 2001) which suggests that there is some sampling error in the PUMS and that perhaps the error is not random. In this study commuters were undercounted with greater undercounting for specific ethnic populations, poor populations (non-home owners), and more so in some MSAs (Metropolitan Statistical Areas) than others, as well as differences across regions of the

country. In general the undercount rate was less than 10%, and more often less than 5%. Apparently, these undercounts are not adjusted for in the sampling weights used with PUMS to represent the whole population.

The US Census (2002) itself points out that only geographic areas from which 200 or more long forms are completed provide adequate data to produce 'good quality estimates', i.e. small confidence intervals. This has been demonstrated to produce distortions in some population area characteristics.

LifeLine used NCHS's Natality data for its population, which is drawn from birth certificates filed in each state. Since 1972 this is a 100% sample in some states and a 50% sample in other states. After 1998, the Natality data is reported to include 99% of hospital births. However, the level of completeness of the data varies by state (NCHS, 1999). For example, six states lacked data on maternal Hispanic origin for more than 5% of births, with more than one-third of all missing data coming from one state. LifeLine did not use paternal race and ethnicity for its Natality indices due to the inadequacy of that data (12 states missing more than 20% of father's race data and 14 states missing more than 20% of father's Hispanic ethnicity data). Four states lacked data on maternal education level [an indicator of SES (Socioeconomic Status)] for 4.5% - 10% of births. In addition, the Natality data does not include children born outside the US and its territories. Therefore, immigrant populations have been excluded.

DEEM-Calendex use of the CSFII population provides much internal consistency since it relies heavily on data obtained from CSFII for reference population, food diary, survey weights and body weights. This extensive reliance on CSFII potentially puts it at risk for whatever inherent biases there are in the CSFII reference population, as well as those associated with telephone surveys.

Several suggestions were made as to further assessments to conduct. In an effort to continuously consider techniques for characterizing the variability in the populations generated, it was suggested that the Agency look into other techniques for generating the populations. One approach may be to consider Bayesian techniques (Gelman et al., 1995).

The development of the population generated by each model appears to be mechanistic, even though they are based on statistical methods (such as sampling based on weights provided in the CSFII data base.) A useful exercise may be to explore other issues of characterizing the population, for example, there may be purely statistical methods for characterizing the population.

Each model is unique in its design for the reference population. Some of the considerations are related to how the two non-consecutive day CSFII diaries can be extended for estimating long-term exposures (e.g., the incompatibility criteria for CARES). Thus, it is understandable that the impact of these model differences may not be apparent in the limited comparison of a single day exposure as presented by the

Agency in "*A Model Comparison: Dietary (Food and Water) Exposure in DEEM/Calendex, CARES, and LifeLine*". It is also noted that the chronic exposures given in Appendix B also did not reveal any significant difference between the exposure estimates from DEEM-FCID and LifeLine. It is recommended that in future model comparisons, the unique design of each model be specifically tested. For example, since CARES is designed to maintain the individual's dietary characteristics (e.g., vegetarian, low fat diet), its impact may not be obvious at a level where all adults within 20 to 49 years old are combined into one population subgroup that includes all seasons, ethnicity, and race.

3. Binning Design & Frequencies of Using CSFII Diaries:

These models differ in the expected (or actual) frequencies that each CSFII diary is used in the probabilistic risk assessment. DEEM-Calendex uses only the individuals that provided two days of food diaries in its reference population, and sets aside approximately 1,000 one-day food diaries in estimating dietary exposure. CARES employs a Gower dissimilarity index in its algorithm to generate longitudinal consumption profiles for its Reference Population. The result is use of some CSFII diaries much more often than other diaries in simulating exposure (as much as 4,000 times for certain diaries versus once for others). Approximately 1,000 CSFII diaries are not included in the CARES Food Match table. The LifeLine model uses a very general bin based on age and season, such that all food diaries within a particular bin have the same expected frequency of being used in its exposure assessment. In order to evaluate the effect of these differing frequencies and modeling weights, EPA approximated all three models using the LifeLine recipes (i.e., keeping recipes constant).

Question 3.1 **The SAP is asked to please comment on the frequency that CSFII diaries are used by the various models. Are there any potential biases that may arise in the respective dietary exposure estimates for these models as a result of how they used CSFII records? Considering LifeLine's current dietary bin design (age, season), please comment with respect to the use of the CSFII survey weight option. Is either LifeLine option (CSFII-weighted or not) generally more appropriate than the other or are there circumstances in which one might be preferable to the other?**

Response

On the Issue of Potential biases: There are several potential biases that can be introduced by the different weighting schemes used by the different models. Biasing calculations one way or the other is, after all, the whole point of using weights in the first place. What the question seems to be asking is whether any of the different approaches is clearly wrong in some way. The answer to that, as far as the Panel can tell, is no. Several

potential sources of bias are discussed below. Although the information presented to the Panel does not suggest that any of the potential biases is necessarily present in any of the models, the Panel does not have the benefit of an exhaustive analysis of the CSFII data and binning methods used by the respective models, and there may be substantial biases that are as yet unrecognized.

The use of CFSII data by all the three probabilistic exposure models could certainly create bias. Perhaps 75% or more of the households have more than one member participating in the CSFII survey. A serious question arises as to whether the food diaries provided by the participants who lived in the same household are related to each other more closely than would be those of random respondents. If this tendency is proven to be true, the binning design used by DEEM/Calendex would have inherited this bias, and the statistical power of model estimation by DEEM, and probably the other two models as well, would be reduced. An autocorrelation analysis for those pair-diary data should be conducted to investigate this issue.

Bias could also arise if the simulated population distribution of one-day food consumption records is not a generally accurate representation of the target population (presumably the United States at some point in time). Because a typical model run may require multiple draws of individual diet records, the potential for bias depends upon the relative frequency with which individual CSFII records are included in a model run (selection and use of population weights notwithstanding). Bias could also arise in other ways. First, bias could result if the binning of diet-days was done in such a manner that systematic and important differences in diet were not an accurate reflection of the demographic attributes used to define a bin. In this case, the weighting of diet records would not be consistent with the true set of diet records, i.e., those for a given demographic sector of the population as defined by the respective reference populations. Bias could also result if sampling procedures from within each diet bin did not yield an equal probability of selection for each observation within the bin.

Potential biases resulting from the way the CSFII records are used could include the following. CARES' exclusion of CSFII diaries may be a problem for the 50-plus age group because approximately 10% of CSFII diaries for this age group were not used. In addition, about 5% of the diaries for the 3–5-year-old group and 5% of the 20–49-year-old group were not used. The CARES materials provided to the Panel did not give an explanation for these exclusions, nor did they comprehensively assess the impact of these exclusions on their model results. Another potential bias involving the impact of the various sampling strategies on the exposure model is evident when an infrequent food source such as olives or papayas is used (page 35). Several Panel members felt that EPA should determine the frequency that individual diet records are selected by the respective models for a standardized modeling scenario. Which diets are highly represented? What characterizes the diets that are never used? This would be a first step to exploring the potential differences, and hence biases, of the population distribution of one-day food consumption records produced by the three models. The population distribution of intake

for each RAC simulated by each model should also be compared to the population distributions in the CSFII.

DEEM/Calendex, LifeLine and CARES each use daily food consumption diary data from the CSFII to model daily food intake for simulated individuals. Past Panels have advocated using multiple donor records from the CSFII to simulate some diversity in individual diets over time. LifeLine and CARES have responded by introducing draws of daily dietary information from stratified pools of CSFII respondents. LifeLine uses a stratified random sampling of a daily dietary record from "bins" that are defined by age and season of the year. The current LifeLine procedure of randomly drawing from the set of available records in each "bin" is a good statistical sampling procedure. Since all records in a bin are available to be sampled at each draw, the procedure is equivalent to sampling with replacement (SWR). A minor improvement to this SWR procedure for selecting dietary records within a bin would be to use sampling without replacement (SWOR). The SWOR procedure could be implemented by first randomly sorting CSFII dietary records (or record pairs) within each bin. The CSFII records in a bin would then be used in random order to simulate the daily diets until all records in the bin have been used once. When the last record in the randomly ordered set of "donors" for a bin has been used, the selection algorithm would re-randomize the dietary records in the bin and then cycle through the ordered list again, continuing in this fashion until the model run is complete. The SWOR sampling of CSFII dietary records ensures that the full statistical information in the CSFII dietary records is used to the maximum extent possible in the simulation. See Cochran (1977) for a discussion of the advantages of SWOR selection over SWR selection.

The procedure used by LifeLine to draw a CSFII daily dietary record for the simulation of daily exposures to pesticides and other chemical compounds is in many ways analogous to the "hot deck" imputation procedure that is used in many large federal survey programs such as the Current Population Survey (CPS) to assign imputed values to missing data items in a survey response. In hot deck imputation, a missing data item (in LifeLine, the daily diet of an individual) is imputed by drawing an observed value from a hot deck cell. Hot deck cells are analogous to the "bins" defined in LifeLine to match simulated individuals with CSFII dietary record donors. In hot deck imputation, the issue of whether to use the survey weights (which reflect population share representation for a case) in the choice of a "donor" value has been debated. This issue of weighted or unweighted choice of donor records has been discussed in the literature, but there is no single opinion or best statistical practice that can be cited. In all practicality, the choice between the two approaches should not make a large difference unless the weights for individual CSFII cases in a bin are highly variable. Guidance and insight on the differing views may be found in the hot deck imputation literature. A good primary reference on hot deck imputation is Little and Rubin (2002).

Finally, more than one Panel member indicated that LifeLine's age-season binning may be a better design, especially if a seasonal effect on food consumption is proven to be a significant factor. One Panel member felt that if stratifying the food diary bins by season

is important, the ultimate model outcomes should not be affected by modeling weight.

Frequency of record use: EPA expressed concern about the fact that some CSFII diaries were used much more often than other diaries in simulating exposures. Several Panel members felt, however, that the mere observation of very large differences in how often diaries get used in the simulations, by itself, is not an indication that there is anything amiss in the CARES model. The sample sizes described for the model runs were relatively small. Perhaps with a different selection of 100,000, a simulation would obtain a different number of missing diets and they would likely be different diets. Even if the observation persists in multiple simulations based on different random seeds, there may not be any real problem. The model could certainly be fashioned so that all the records were used, but it was a modeling decision not to do this, and it is not clear that reversing this decision would represent an improvement. The different frequencies of use are the result of the model's taking account of the correlation structure between individuals and dietary patterns, which is a very important feature of the program. Removing this structure, such as by selecting diets at random, would destroy this structure.

There may, nevertheless, be room to compromise by relaxing the specification of the Gower dissimilarity index. [The Gower dissimilarity index (Gower 1971), a measure of the distance between two multivariate objects, is similar to the Manhattan metric.] In principle, the proportion of diaries that are potentially randomly selected in the simulations used could be adjusted by altering the threshold of dissimilarity that defines when diet records match. However, doing so might require EPA to enlist the cooperation of the model programmers. CARES' use of the Gower dissimilarity index could allow each age group to sample from a large group of CSFII diets. Using the Gower dissimilarity index may include more diaries than DEEM, and perhaps be similar to LifeLine.

The binning design associated with CARES is directly related to how the Gower dissimilarity indices are set. The total number of food diaries in a particular age bin (such as 50-plus age) that are not used by the model may not have anything to do with the demographic information of this age group, as speculated by EPA, but rather is determined by how the constraints are set by the Gower dissimilarity index. One Panel member was interested in the observation that adjacent age groups sometimes had dramatically different numbers of food diaries that were not being used (Table 7, page 22). For example, there are 394 food diaries in the age 3–5 bin that were not used by CARES, however, all the food diaries in the next age group bin were used. Were the foods consumed by these two age groups different? Were the individuals different between CSFII and the Gower dissimilarity index?

Finally, it was noted that when the CSFII diaries show infrequent events, random selection of diaries may either exclude or include these infrequent events. The random inclusion of these infrequent events can and does have a major influence on model outcomes.

Model uncertainty: In the model comparisons undertaken by EPA, some of the observed differences were traced to glitches in the data entry or transformation steps. Presumably such glitches will be fairly rare in routine use of these models. In other cases, when there is a difference that is really due to the way the different models work, it may not be clear that one model is better or best. In such cases, there is genuine model uncertainty. This model uncertainty should be respected and propagated into regulatory decisions. The Panel recognizes that decision makers don't like it when the model disagreement "spans regions of regulatory interest", but this is just exactly when it is important to acknowledge the limits in our ability to forecast.

The differences among the models are not the consequence of any advanced or subtle mathematics. If they were, there could be an interesting debate about what method would be best. All the models presented are accounting programs, in various implementations. They have different ways to do the accounting, but their differences may be pretty close to being "six of one versus half a dozen of the other", at least with respect to what is known empirically about aggregate dietary exposures, where the science is. Given the state of the science (whatever the dietary and exposure experts say it is), it may be that the only way to discern whether a model is biased is to compare its outputs to the real world. The Panel understands that to compare these results to real situations requires collection of extensive diets from a large number of individuals, an expensive and time consuming task. Despite this, the Panel strongly encourages EPA to get these data now and not defer this task to some unspecified future time.

Similarity of model results: The numerical comparisons of the models described by EPA revealed very little difference among them. For instance, it seems clear to the Panel that the overall uncertainty arising from limited sample sizes in the input data is probably far greater by itself than some of the differences observed in the model comparison exercises. Several members of the Panel noted that the model comparison exercises seemed to minimize the differences between the models. This was presumably intentional, because EPA was looking for mistakes or peculiarities in the implementations of the models. Nevertheless, barring any gross inconsistencies arising from programming errors or oddities in the underlying conceptual models, the disagreement among the models is itself of much interest because it represents part of the model uncertainty in the exposure assessment. The Panel feels that it is still unclear how different the model results could be in interesting cases. EPA should continue its work to address how large the discrepancies might be.

In particular, several Panel members noted that when comparing results from a single-day, simply using identical inputs is not a sufficient test for the process. Of more interest than the single-day exposure is the evaluation of longitudinal exposures for various populations. This is what the models are designed to do, and eventually they must be put to this test. Of particular interest is to follow a cohort, e.g., the one used in LifeLine, and to develop an assessment of long-term trends. Important in this process is evaluating potential changes in a population's eating habits. Examples of such changes would be

the trends in the 1980s and 1990s toward diets lower in fat, and the rise in use of bottled water. The use of different populations, all selected for good scientific reasons, aids in our understanding of the influence of populations and changes that are likely to occur in dietary exposure in future years.

Multiple models: It would clearly be wrong and premature to infer from the comparisons conducted so far that it would be okay to choose one of the three (or four) models and dispense with the other models. (This includes choosing one CSFII-weighting for LifeLine.) The Panel believes that EPA needs to continue to run several model/options to explore the model uncertainty. Indeed, it would generally be best to employ as many models as are available, because each model is likely to have strengths and weaknesses that would be offset by others. One model may address parts of the dietary exposure better than another, while others would improve on different areas. The knowledge that can be inferred by combining results from all of them would take advantage of an economy of scale, the whole being greater than the sum of the parts. One Panel member suggested that the EPA should continue to develop not only these models, but also include simple statistical and mechanistic models. Such models may be easier to build, require less obscure input information, and provide a considerably more transparent analysis. This may help EPA to arrive at faster, more cost-effective results. The development of the SAS models is perhaps an example of this kind of approach. EPA should be accepting of models that may not "do everything" but will provide simple solutions for specific objectives.

4. Commodity Exposure Contribution Analyses:

An important aspect of any dietary risk assessment is the ability to identify significant contributors at the upper percentiles of exposure. The CARES and DEEM models both include an output report option known as the Critical Exposure Contribution (CEC) analysis. A comparable report option is expected to be developed for the LifeLine model in the near future. These CEC reports quantify the contribution of specific food commodities (RAC-FF) to the total exposure at the upper percentiles (e.g., top 0.2%) of the exposure distribution. An alternate or complementary approach (frequency-exceeded), also used by various model developers, tabulates the frequency that a particular commodity (RAC-FF) causes exposure to exceed some level of concern. As was the case with predictive exposure estimates, model design can affect the outcome of commodity exposure contribution analyses. Section IV.G of the background document describes the CEC and 'frequency-exceeded' approaches for identifying significant contributors at the upper end of the exposure distribution. Tables 13 and 14 show CEC reports and 'frequency of occurrence' data for DEEM-FCID and CARES analyses for 3 - 5 year olds and 20 - 49 year olds, respectively. Tables 15 and 16 show SAS approximations for the model CEC reports and 'number of occurrences > aPAD' for these same age groups. Although there is certainly a degree of similarity between model results and between the model results and the SAS approximation results, differences do occur.

Question 4.1 **The SAP is asked to please comment on the relative merits of the two approaches described above (CEC and frequency-exceeded) for identifying significant contributors (RAC-FF) to exposure at the upper percentiles of exposure. Are there other methods or techniques which the Panel might recommend for accomplishing this important part of the dietary exposure assessment?**

Response

The Panel recognizes that the Critical Exposure Commodities (CEC) is a group comparison tool while the "frequency-exceeded" measures are self-references. Both measures contain arbitrary components (i.e., "top 0.2%" and "some level of concern") that are incomparable and are associated with different cutoff values. Both CEC and "frequency-exceeded" analysis are valuable for identifying significant contributors to the dietary exposure. They are practical measures for characterizing the overall risk and useful for highlighting potential areas for further refinement of the exposure analysis as well as risk mitigation. Both measures should be used to ensure that a RAC with a high level of risk would not "slip through the system".

It was noted that the Agency's SAS method can both facilitate the understanding of different model outputs and be used for a quick estimate of exposure scenarios that might exceed the Population Adjusted Dose (PAD). Until an all-inclusive measure for risk management is identified, all methods of measures should be explored.

The Panel suggests further investigations into the following areas of dietary exposure analysis. Besides the bigger picture of population-wide exposure analysis, it is important that the single-day exposure analysis also identifies individuals at a reasonable high end of exposure (e.g., upper 0.1 percentile) to ensure that their exposure would not exceed the acute PAD (aPAD). In addition to the CEC and the "frequency-exceeded" analyses, the high end exposure can be quickly identified by making use of the CSFII consumption profiles for the 400 plus RACs and their associated food forms now developed by the Agency. The Agency is encouraged to compile and publish these consumption records. At the start of a risk assessment, the potential PAD exceeders can then be identified by simply multiplying the high end residue level with the high end consumption rate. Along the same line, the Agency is further encouraged to look into identifying specific patterns of potential higher exposure (e.g., seasonal pattern). Iterative assessment can also be conducted to examine the impact of one food by omitting it from the subsequent re-analysis.

In addition, the Agency should further investigate the impact of CSFII diaries that are not used by the models. While some explanation was given for not using the 1-day only diaries in DEEM-FCID model, the characteristics of the unmatched and therefore unused

diaries by CARES are largely unknown. With the preciously limited size of the CSFII data, it is of interest to find out about these unmatched records and see if they contain essential or unique dietary characteristics and records that are currently not accounted for.

Distinction can also be made of the two types of RAC-residue pairs that are important for exceeding thresholds of concern (e.g., Population Adjusted Dose, PAD). One is foods with high residue concentration that are large contributors to one-day exposure when eaten. Another is foods with more modest residue concentrations but higher consumption rates on a population basis (i.e., staple foods). Recognizing the contributions of both types of RAC-residues is important for making protective and efficient tolerance decisions.

The ultimate challenge for a single-day exposure analysis is to identify and protect against the risk of a reasonable range of high-end exposure of a single RAC or RAC-FF, as well as the total exposure from multiple RACs. The assessment should address exposures at both the individual (e.g., individuals having reasonably high single day consumption) and the population levels (e.g., children subgroups that have generally higher exposures).

REFERENCES

Cochran, W.G. 1977. Sampling Techniques. 3rd Edition. John Wiley and Sons, NewYork.

Gelman, A., J.B. Carlin, H.S. Stern, D.B. Rubin. 1995. Bayesian Data Analysis. Chapman & Hall/CRC Press. Boca Raton, FL.

Gower, J.C. 1971. A general coefficient of similarity and some of its properties. Biometrics 27:857-871.

Lin, Chia-yu, A. Gelman, P.N. Price, D.H. Krantz. 1999. Analysis of Local Decisions Using Hierarchical Modeling, Applied to Home Radon Measurement and Remediation. Statistical Sciences 14(3):305-337.

Little, R.J.A. and D.B. Rubin. 2002. Statistical Analysis with Missing Data, 2nd Edition. John Wiley and Sons, New York.

National Center for Health Statistics. Technical appendix. Vital statistics of the United States: Natality. U.S. Dept of Health and Human Services, Centers for Disease Control

and Prevention. **http://www.cdc.gov/nchs/data/technap99.pdf**

Ong, P. and E. Ong. 2001. Undercounting commuters: report to US census monitoring board. February 2001. Report # 07.

US Census Bureau. August 9, 2002. Comparing SF 3 estimates with corresponding values in SF1 and SF2. **http://www.census.gov/Press-Release/www/2002/sf3compnote.html**