UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

WASHINGTON, D.C. 20460

OFFICE OF
PREVENTION, PESTICIDES
AND TOXIC SUBSTANCES

March 31, 2004

MEMORANDUM

Subject: Transmission of Background Materials and Charge to the Panel for the Session of the April 29-30, 2004 FIFRA Scientific Advisory Panel Entitled "A Model Comparison: Dietary and Aggregate Exposure in Calendex, CARES, and Lifeline"

To: Myrta Christian, Designated Federal Official
FIFRA SAP Office of Science Coordination and Policy (7101C)
Francis Suhre, Branch Chief

From: Office of Pesticide Programs
Health Effects Division
(7509C)
Margaret Stasikowski, Director

Through: Office of Pesticide Programs,
Health Effects Division
(7509C)

Attached is the document "A Model Comparison: Dietary (Food and Water) Exposure in DEEM/Calendex, CARES, and LifeLine." This document provides background information to the SAP on each of the three models, compares exposure estimates and results generated by each of the three models using a common data set for a hypothetical chemical, and investigates specific reasons why-- and to what extent -- exposure estimates generated by the three models might be expected to differ. Such differences are of interest to OPP in a regulatory context, particularly if the differences are large, are systematic, or span regions of regulatory interest.

The current SAP session for which this material is being provided is a "follow-on" to previous SAP sessions in which each model was reviewed by the Panel. At that

time, the SAP provided valuable feedback to the Agency. The previous material from these past SAP sessions is again being provided to the Panel. A list of this material is attached to this transmittal letter.

The following material is provided to the Panel members:

A Model Comparison: Dietary (Food and Water) Exposure in DEEM/Calendex, CARES, and LifeLine. document dated March 31, 2004

CD- ROM containing material indicated in attached document entitled "Background Materials for FIFRA Scientific Advisory Panel Meeting"

****

Charge to the Panel:
Please provide comment and advice on the following issues and questions:

Question 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
  - 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.

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.

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.

Question 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.

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

Question 3. <u>Binning Design & Frequencies of using CSFII diaries</u>:

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).

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?

Question 4. <u>Commodity Exposure Contribution Analyses:</u>

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.

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?

Attachments: List of Background Material on CD-ROM provided to FIFRA SAP Members

Background Materials for FIFRA Scientific Advisory Panel Meeting

April 29-30, 2004

(on CD-ROM)

A Model Comparison: Dietary (Food and Water) Exposure in DEEM/Calendex, CARES, and LifeLine. document dated March 31, 2004

PREVIOUS SAP MATERIAL (listed by exposure model)

Calendex

FIFRA Scientific Advisory Panel Meeting: September 27, 2000.
Calendex. Calendar-Based Dietary & Non-dietary Aggregate and Cumulative Exposure Software System. 124 pp.

SAP Report No. 2000-03. FIFRA Scientific Advisory Panel Meeting: September 27-29, 2000 *Partial Report Residential Exposure Model-Rex, Calendex and Aggregate and Cumulative Assessments using Lifeline.* [see Section IV -Calendex Model] 32 pp.

CARES

SAP Meeting No, 2002-02. FIFRA Scientific Advisory Panel Meeting: April 30-May 1, 2002 *Cumulative and Aggregate Risk Evaluation System (CARES) Model Review*. 27 pp.

CARES v 1.0 *Cumulative and Aggregate Risk Evaluation System Code Manual*. 100 pp.
March 20, 2002

CARES *Appendix A Glossary and Definition*

CARES v 1.0 *Cumulative and Aggregate Risk Evaluation Technical Manual* 3/20/02. 172 pp.

CARES v 1.0 *Cumulative and Aggregate Risk Evaluation User Guide* 3/20/02. 222 pp.

DEEM

SAP Meeting. FIFRA Scientific Advisory Panel Meeting: February 29- March 3, 2000 *Dietary Exposure Evaluation Model (DEEM) and DEEM Decompositing Procedure and Software.* 146 pp.

SAP Report No. 2000-01. FIFRA Scientific Advisory Panel Meeting: February 29- March 3, 2000 *Partial Report Dietary Exposure Evaluation Model (DEEM), MaxLIP and DEEM Decompositing Procedure and Software.* [see Section IV -Calendex Model] 56 pp.

Lifeline

Lifeline March 2001 SAP

SAP Report No. 2001-07. FIFRA Scientific Advisory Panel Meeting: March 28, 2001 *HRI Lifeline- System Operation Review*. 23 pp.

Lifeline v. 1.0 *Aggregate and Cumulative Exposures to Pesticide User Manual* December 15, 2000. 138 pp.

Lifeline Sept 2000 SAP

Overview of the Fundamentals of Version 1.0 of *Lifeline Software for Modeling Aggregate and Cumulative Exposures to Pesticides* September 2002. 9 pp.

SAP Background Document. FIFRA Scientific Advisory Panel Meeting: September 1999 *Review of an Aggregate Exposure Assessment Tool*. 23 pp.

Introduction and History of Lifeline. C.F. Chaisson August 25, 2000. 5 pp.

Assessing Aggregate and Cumulative Pesticide Risks Using Lifeline v1.0. August 31, 2000. 26 pp.

Lifeline Sept 1999 SAP

SAP Meeting Background Documentations. FIFRA Scientific Advisory Panel Meeting: September 1999 *Review of an Aggregate Exposure Assessment Tool*

SAP Report No. 99-05. FIFRA Scientific Advisory Panel Meeting: September 21-24, 1999 *Review of an Aggregate Exposure Assessment Tool*. [see Session II] 75 pp.

CURRENT USERS/TECHNICAL MANUALS

Lifeline:

Lifeline v. 2.0 *Aggregate and Cumulative Exposures to Pesticide Technical Manual* July 31, 2001.

DEEM:

DEEM v. 7.0 *Dietary Exposure Evaluation Model Users Manual.* May 2000.

CARES:

CARES v 1.0 *Cumulative and Aggregate Risk Evaluation Technical Manual* 3/20/02. 172 pp.

CARES v 1.0 *Cumulative and Aggregate Risk Evaluation User Guide* 3/20/02. 222 pp.

[*these latter two CARES documents are contained in the CARES SAP folder as they were presented to April 30/May 1 2002 SAP*]

SAS MATERIAL

SAS Programs (*.SAS)

DEEMCalendex_CONSAPPROX.SAS - The SAS program used to construct the approximations for both models (D1, D2).

CARES_CONSAPPROX.SAS - The SAS program used to construct the approximations for CARES (C1). The program CARES_AccessImport_RefPop.SAS is used to tabulate the weighted frequencies that each CSFII diary was used in the CARES' Food Match tables (Jan-Dec).

Lifeline_CONSAPPROX.SAS - The SAS program used to construct the approximations for Lifeline (L1).

ModelComparison_CEC - contains several macros: (i) %LL_ImportResidues(datain) imports residues from Lifeline (Food Residue Translator - Export Inputs) into SAS, (ii) %LL_AcuteCEC merges residues with consumption distributions, calculates exposure - and corresponding probabilities for all permutations, and the expected distribution for exposure, by commodity (CG-RAC-FF), (iii) %Exceeders calculates the number of consumption-residue permutations that lead to exposure exceeding the aPAD, (iv) %CALC_CEC

calculates exposure at the 99.9th percentiles, as well as the shares of total exposure for the top 0.2% for each commodity (CG-RAC-FF), (v) %Compare5 merges the consumption distributions to help identify commodities where the models are predicted to differ, and (vi) %PlotRAC plots the consumption distributions for each of the model approximations.

SAS Model Approximations (*.sas7bdat)

DEEM-Calendex (CONS_*AgeGroup*_D1) - approximates model using only the 2 day diaries and corresponding survey weights (WT4_2DAY) for each age group.

DEEM-Calendex (CONS_*AgeGroup*_D2) - approximates model with the hypothetical situation assuming that all food diaries and corresponding survey weights (WT4_DAY1) were used by DEEM-Calendex.

CARES (CONS_*AgeGroup*_C1) - approximates model for each age group. The model weights (Jan_Dec_*AgeGroup*) account for both the frequencies that the corresponding diaries were used and the model weights applied for each use.

Lifeline (CONS_*AgeGroup*_L1) - approximates model for each age group. The model weights (LL1_Dietbin_Count_*Subgroup*) account for both the frequencies that the corresponding diaries were used (inversely related to bin sizes) and the model weights (=1) applied for each use. The total projected population (Total_WT4Day1_LL1_*Subgroup*) assumes that the mortality option is not used (i.e., everyone lives to age 85).

Lifeline (CONS_*AgeGroup*_L2) - approximates the Lifeline model with the hypothetical assumption that the CSFII respondents' reported body weights are used (rather than the Lifeline anthropometric modeled body weights). Since this model does not require generating permutations of CSFII diaries and Lifeline body weights, the total projected populations (Total_WT4Day1_LL2_*Subgroup*) are equal to the total number of person-seasons in each subgroup, e.g., equal to 4 for infants (=1 yr x 4 seasons/yr), 8 for 1to2 year olds (=2 yrs x 4 seasons/yr), and so on.

Lifeline (CONS_*AgeGroup*_L3) - approximates the Lifeline model with the assumption that the CSFII survey weight option is used to affect the probabilities of selecting among food diaries within a dietary bin. For each dietary bin, these probabilities (LL3_Dietbin_Count_*Subgroup*) are directly proportional to the CSFII survey weights.

CSFII demographic (RT25_LL) and food diary (RT30_LL) tables contain the food survey data.

Lifeline Food Recipe (Rtrecp11_LL) lists all of the ingredients and corresponding percentages (mass) for all food items.

Food residues for the simple 1,000 ppm examples (Residue_SingleRAC) and the MultiRAC case study (Residue_MultiRAC) that can be matched with the list of Lifeline CG-RAC-FF.

SAS Output Tables (*.sas7bdat)

Comparisons for Simple 1,000 ppm examples across the 5 models (D1,C1,L1,L2,L3) (Z5_AgeGroup_dcl1l2l3), by commodity.

Results (%LL_AcuteCEC), by commodity, for the Simple 1,000 ppm examples (SIMA_SingleRAC_*AgeGroup_Model*), as well as the MultiRAC case study (SIMA_MultiRAC_*AgeGroup_Model*). The CEC reports (%CALC_CEC) (CEC_SIMA_MultiRAC_*AgeGroup_Model*), and the exceedence report (%Exceeders), for the MultiRAC case study, by commodity.