Background Document for the Session:

Review of an Aggregate Exposure Assessment Tool

Presented by:

Paul S. Price, M.S.
(Ogden Environmental and Energy Services)
Christine F. Chaisson, Ph.D.
(CF Chaisson Scientific Advisors)
John S. Young, Ph.D.
(The Hampshire Research Institute)

Carol Christensen
Elizabeth Doyle, Ph.D.
Francis B. Suhre, M.S.

United States Environmental Protection Agency
Office of Pesticide Programs
401 M St., S.W.
Washington, D.C. 20460

to the:

**FIFRA Scientific Advisory Panel,
Arlington, Virginia
September, 1999**

THE LIFELINE™ PROJECT TO MODEL AGGREGATE EXPOSURES TO PESTICIDES

Project Description

The LifeLine™ project is an effort to develop open, verifiable, publicly available software to perform aggregate (*i.e.*, multiple sources and routes of exposure) and cumulative (*i.e.*, exposure to multiple chemicals with a common mechanism of toxic action) exposure and risk assessment. This paper presents a brief description of the modeling approach employed and discusses key scientific and technical issues in that approach. The LifeLine™ project continues the Hampshire Research Institute's core program of providing fully open risk exposure and risk assessment software to the public on a basis that provides only for the recovery of ongoing technical support costs, so that users need not support development costs. The initial version, supported by EPA, USDA, and private funds, focuses on residential, dietary and tap water exposures to pesticides.

Background

For much of its history, the field of exposure assessment has focused on characterizing the highest levels of exposure that will occur to an individual or a population over time as the result of the use of a pesticide. One approach that is used to characterize the upper bound of exposure is to use simple models of dose rates and a series of conservative model inputs. This approach has great value for screening-out exposures that are of little concern. A related approach is to back off from one or more of the "worst-case" assumptions and use a mixture of conservative and more reasonable estimates. This approach increases the confidence that the identified exposures represent actual risks. These two approaches form the basis for EPA exposure guidance such as *Risk Assessment Guidance for Superfund* (RAGS; EPA, 1988) and the draft Residential SOPs.

The difficulty with these approaches is that an individual who may be receiving high levels of exposure from one source will not necessarily receive high levels of exposure from a second or a third source. In fact, there are situations in which exposure to high levels from one source will preclude exposure from a second source. As a result, exposure assessment approaches that focus on defining individuals who have high levels of exposure to a single source cannot be extended to evaluate multiple sources. What is needed is an approach that tracks the simultaneous exposure to multiple sources.

Two solutions have been suggested for this problem. The first is to collect data on the simultaneous exposures of individuals from all sources of a pesticide in each of the individuals' life. This requires surveying all behaviors that are important to defining an individual's exposures to each of multiple sources. This approach is currently used in dietary exposure software that uses daily dietary records to evaluate simultaneous exposure to pesticide residues that occur from the consumption of multiple agricultural commodities.[1] A similar approach is used in several air exposure models. This approach has the advantage of capturing the correlations between the individual's actions. Thus, a single survey record could accurately determine the inputs to dose rate models for multiple concurrent sources.

There are, however, severe drawbacks to this approach. First, it is difficult to obtain survey results on an individual's behaviors (either food consumption or activity patterns) for periods

---

[1]   It should be noted that even these models combine data from multiple surveys (residue surveys, market basket surveys, as well as dietary surveys).

longer than one or two days. Therefore, a single survey-based approach cannot be used to accurately evaluate exposures that occur over longer periods of time, or that occur infrequently (*e.g*., use of termiticides). In addition, as the number of potential sources increases, the number of behaviors that must be investigated in a survey increases proportionately. Such an approach cannot be applied to the problem of aggregate exposure assessment, absent a major effort to collect data on individual behaviors with a thoroughness never before attempted.

The alternative approach is to simulate the total dose received from multiple sources by individuals in a population. Monte Carlo analysis is often used for these simulations (McKone and Ryan, 1989; McKone and Daniels, 1991). These models have allowed the incorporation of data from multiple surveys. Examples of such models include CalTOX and MCCEM.

Monte Carlo analysis is equally applicable for simple or complex dose rate models (Morgan and Henrion, 1990). The application of Monte Carlo to complex time-dependent exposure models is called Microexposure Event Analysis (Price et al. 1992, 1996; Keenan et al. 1993; Harrington et al.1995; Goodrum et al. 1996). The technique of Microexposure Event Analysis has been proposed for use in evaluating both aggregate and cumulative exposure (Muir *et al.*, 1998).

The software being developed in the LifeLine^{TM} project represents such a simulation of exposure and draws on data from a number of different surveys. Information on daily activity and dietary patterns from well-known surveys is used to evaluate specific daily exposures for an individual. Data on demographics, residential pesticide uses, and residential characteristics are being drawn from multiple surveys. The software uses the most appropriate database to address each component of the simulation. Data sets that are used in the initial version of the model include:

- Natality data (Birth records databases maintained by the National Center for Health Statistics [NCHS])
- Residential mobility (US Census)
- The Third National Health and Nutrition Examination Survey (NHANES III), also maintained by NCHS
- American Housing Survey (Census)
- Nation Home and Garden Pesticide Use Survey (EPA)
- National Human Activity Pattern Survey (EPA)
- The Continuing Survey of Food Intake by Individuals (CSFII), from the Department of Agriculture (USDA)
- Residential Exposure SOPs (EPA)
- *Exposure Factors Handbook* (EPA)

Future versions of the model will also incorporate additional data on pesticide residues:

- Pesticide Data Program (PDP) from USDA
- Field Trials Data (EPA)
- Surveillance Monitoring Data from the Food and Drug Administration (FDA)

These combined data sets allow the model to define the exposure for each day of an individual's life. The model does this by modeling where people are born, how individuals grow and age,

how they move from home to home and region to region, how they use or do not use pesticides, and what are their daily activity and dietary patterns.

Performing this assessment require careful planning and organization of data. This organization was achieved based on a series of modeling principles called "transition rules". These rules are discussed in the following section.

Modeling Time Varying Exposures Using Transition Rules
Introduction

The LifeLine™ Project software is a Monte Carlo (probabilistic) model of the aggregate exposure to pesticides that occurs to each member of a specific population of individuals. The key focus of the software is modeling each potentially exposed individual within that population *as an individual*. Specifically, the model seeks to define each simulation of an individual in such a way as to provide an accurate characterization of inter-individual differences in exposure-related behaviors for populations of interest. This simulation must assign all of the individual's characteristics in an internally consistent way and in a manner that reflects the population under investigation.

This approach addresses the criticism that Monte Carlo models of inter-individual variation in dose create "implausible" or chimeric individuals. These individuals combine values for physiological characteristics or behaviors that were not internally consistent. Examples of such inconsistencies might include a large surface area and a small body weight, an assumption that a person performed an implausible number of activities on a single day, or the use of an air exchange rate from a home in Maine and a pesticide use pattern for a home in Florida.

Modeling time-varying exposures compounds the task of developing internally consistent values for inputs since each day (or partial day) is modeled separately. Such models must assign tens of thousands of values for each input of an exposure equation (one for each day of an individual's life). These inputs must be consistent with both the values given to other inputs and to the values given at other times to the same input. The assignment of input values requires careful consideration of whether the value of an input to the model varies over multiple days. If the input's value does vary, the model must accurately capture this variation.

The LifeLine™ Project software addresses this problem by a set of modeling principles called transition rules. These rules specify how a value of an input is initially selected based on the characteristics of the population to be modeled and when and how the input values change over time. An explicit elaboration of transition rules (or equivalent concepts) is fundamental to evaluating any model that makes use of multiple data sources to simulate an individual or population over time.

Classifying Modeling Inputs

One of the transition rules used in the LifeLine™ Project software states that model inputs should be divided into one or more categories depending on their temporal characteristics. The categories are fixed, long-term trends, episodic, cyclic, and ephemeral. In general, the values to the inputs assigning in a specific order:

- fixed inputs,
- properties that vary slowly over time (long term trends, episodic or cyclic inputs), and
- properties that vary from day to day (ephemeral inputs).

Whenever properties that vary over time are changed, the new values must be consistent with prior values assigned to an individual.

Inputs with Fixed Values

An individual has certain characteristics that are constant over her or his lifetime. These include sex, race, ethnicity, birth date, body type, and certain other physiological characteristics. In the model, these inputs are assigned at birth based on the distribution of the inputs in the population of interest (*e.g.*, total US population, blacks, low-income, *etc.*)

The major source of data on these fixed properties is the nation's birth records. The National Center for Health Statistics annually collects data on the nation's births and publishes the data in its Natality surveys.

The software selects a birth record based on the current birth data for the US population or any subpopulations specified by the user. This record provides the characteristics of the individual's sex, race, ethnicity, and place of birth and mother's residence (not only region but also urbanization). The advantage of this approach is that use of a single record to define these inputs will automatically account for the correlations between the inputs selected.

Time-Varying Inputs

Once an individual has been assigned permanent characteristics, an individual's time-varying inputs are assigned for each day of the individual's life. The assignment of the values on any one day is contingent on the values assigned to prior days. Thus, the model begins with exposure during an individual's childhood. For example, the characteristics of the individual's first home is based on the region and setting (urban or rural) of the mother's home, as well as maternal socioeconomic status (SES). The diet and activity patterns are taken from records of infants.

The values of the inputs to the dose rate models vary at different rates in different manners. The inputs can be classified in the following categories.

Long-Term Progressions

These inputs include most of an individual's physiological characteristics. Such characteristics vary in predictable patterns. Height increases regularly until adulthood, and remains relatively constant thereafter.[2] Weights similarly increase with height until adulthood. All of these inputs are influenced by and are thus contingent on the fixed properties (sex, race, and ethnicity). Therefore, the selection of the values are correlated with the characteristics already assigned to the individual.

---

[2]   Slight decreases in height in the elderly are documented as a result of pathological processes; these are not addressed in the model.

In general, age-specific data on these inputs are taken from the findings of the NHANES III survey of physiological data. Since the values certain physiological characteristics are correlated, the values must be selected in a specific order. The height is first modeled,[3] and then contingent on the height, the body weight of the individual is modeled. Based on these two inputs, the surface area of the total body and specific body parts are determined.

Other inputs that follow long term trends may include levels of residues in tapwater or indoor air emissions from termiticides. The user enters information on the levels and temporal trends of these residues.

Episodic Changes (Non-Periodic State Changes)[4]

Many inputs change in an episodic fashion, that is they remain constant for lengthy periods of time, change radically, and then remain constant an additional period of time. Inputs that fall in this category include; residence-related inputs (room sizes, pest pressures, or the presence of a pool or garden), occupationally related inputs, and inputs related to exposures in institutional settings (*e.g.*, school, college, or armed services).

Episodic changes are modeled in a different fashion than other inputs. In general, at the end of each day (or some other suitable period), the model uses a simple binomial decision on whether these inputs change or not. If the inputs do change (the person moves, takes up a new job, enters school, or begins school) then new values are adopted. The approach makes use of age-specific probability of episodic changes that are available from the US Census.

The model places considerable emphasis on residential mobility. The Current Population Survey and American Housing Survey maintained by the US Census provide detailed information on residential mobility. These data are used to determine the probability that an individual of a certain age, gender, and SES, living in a specific setting (urban or rural), region, and housing type (single or multiple family residence) will stay in the same location, move to a comparable dwelling unit, or move to a different region, setting, and housing type.

Mortality is modeled as the ultimate episodic change. The model determines[5] mortality annually, based on the probability of dying from any cause. The probability used is based on age, race, and ethnicity, reflecting life tables published by the National Center for Health Statistics.[6]

---

[3]   Because NHANES III data are not sufficiently rich to support non-parametric selection of heights and weights (*e.g.*, the 10th percentile weight within the 70th percentile height), parametric models including terms for random variability in each correlation are employed.

[4]   Our usage of this term to describe a class of model inputs should be distinguished from historic use of the term "episode" to describe brief periods of exposure to pesticides, whether unique or recurring.

[5]   For policy purposes, the user has the option to ignore mortality and directly specify the lifespan for all individuals.

[6]   There are published problems with some of these tables, which have been addressed as well as possible.

Cyclic Inputs (Periodic State Changes)

Many inputs are determined or are influenced by the season and day of the week. These inputs include activity patterns (weekend versus weekday), diet, residential pesticide use, and dietary residues. The model tracks each day of the individual's exposures as occurring on a specific day of the week and in a specific season. This tracking is based on the birthdate (month and day) of the individual and the number of days modeled. The data on the season and day of the week are used in selecting records of dietary consumption (CSFII) and daily activity records (NHAPS). Therefore, the activity pattern and the diet will be sampled from the correct season and day of the week.

Seasonal variation of residues will also fall into this category. Since many foods come from different regions at different times of the year the pesticides use on a food will change with the season. The user will be able to vary the presence and levels of pesticide residues in each Raw Agricultural Commodity (RAC) for each season of the year. Finally, tapwater residues for surface water systems have also been shown to have seasonal changes. Such data can be incorporated into the model.

In future versions of the software, the exposures related to school, occupational, and recreational environments will be modeled.

Ephemeral Inputs (Vary from Day-to-Day)

Ephemeral inputs are those inputs that vary from day to day. They are individual incidents that may occur more than one time, but without predictable periodicity. Some examples include:

- Was a fogger used in the living room of this house?
- How much time did the individual spend in each area (*e.g.*, each room, as well as lawn/garden)?
- What activity was performed in each room?
- What did the individual eat that day?
- What agricultural commodities were represented in the individual's diet, and what were the sources of each?
- What were the residues in each commodity?

These inputs may be purely random. For example, the residue on a food item purchased from a store can only be modeled as a random sample from the appropriate residue distribution with currently available and anticipated data. These factors are relatively simple to model.

Other ephemeral inputs have aspects that are a mixture of random and cyclical or episodic behaviors. These inputs are modeled as constrained random models. For example, a pesticide used in a residence may be random in the sense that the pesticide is equally likely to be used on a Tuesday or a Wednesday. The probability of the use, however, is influenced by a large number of inputs such as the season of the year, region of the country, type of home, frequency of use, and time since last use.

Transition Rules

In the LifeLine™ Project software, the following rules are used to evaluate the temporal changes in input values in a simulation of a hypothetical individual's life.

1. Inputs are defined in terms of one or more of the following categories: fixed, long-term trends, episodic, cyclic, and ephemeral.
2. An individual's fixed characteristics are always assigned first. This allows the fixed characteristics to be used in the consistent selection of subsequent variables. Values selected should be internally consistent.
3. Each day of a person's life is defined in terms of season and whether it is a weekend or weekday.
4. An individual's time-varying behaviors are assigned starting at birth (or at the earliest age of interest) and proceed through time. Values on any day are consistent with the values assigned to prior days.
5. Temporal changes in episodic variables are modeled by a series of binomial decisions (the variable either changes or remains the same.) The decision is made on a daily basis (or at some other appropriate frequency). The probability of change and the selection of new values is determined from studies of populations that are consistent with the individual's age and other characteristics previously assigned. Once a change has been made (change in residence, etc.), all affected variables, but only affected variables, are modified. These binomial models are in some ways similar to branching models but are more flexible and have the advantage of not requiring the user to exhaustively define all possible outcomes.
6. Selection of ephemeral inputs is based on a random or constrained random model. These models make take several forms. One method is to randomly sample from records that are constrained to be consistent with relevant inputs such as the day of the week, season, age of the individual, gender, residence type, and region. This approach is used for selecting activity patterns and dietary records. A second method is to use a binomial model where the probability of an input changing is contingent on relevant inputs such as season, region, prior use, and residence. This approach is used in the modeling residential pesticide use.
7. The temporal patterns of change for inputs are determined independently. Changes in input values are never automatically linked, unless there is a sound reason for predicting a correlation. Thus, moving to a new home does not change an individual's height but does change the frequency of use of residential pesticides.
8. Where the available data are insufficient to separate inter- and intra-individual variation in an input, the software should allows the user to investigate the impact of assuming that the variation is entirely inter-individual or entirely intra-individual. This rule follows the observation of Buck et al. (1995), that short-term measurements can be viewed as the combination of long-term inter-individual variation and short-term temporal variation.

Dealing with Correlations Between Inputs

The transition rules used in this approach address inter- and intra-input correlations in a number of ways. Two major strategies are the use of records wherever possible and the selection of values that are contingent upon the individual's established characteristics.

By records we mean data sets where multiple inputs of interest are collected at the same time for the same individual (*i.e.*, from the same database record). Examples of such records include the daily activity patterns (NHAPS), the dietary records (CSFII), Natality records, Census records for mobility and housing, and residential use of pesticides (NHGPUS). The selection of values for multiple inputs from such records automatically incorporates the correlations between the inputs.

The second approach is the use of contingent modeling. Under this approach, data are organized into a series of contingency tables that are used to guide the selection of input values. An example of this approach is the selection of height, weight, and surface area. Height is tracked across an individual's life. At each age, an individual's height is determined based on the individual's height at an earlier age and the individual's sex, race, ethnicity, and age. Given this new height, a new weight is selected based on the individual's height. Once the height and weight are selected, the total surface area and the surface area for the hands and other body parts are selected. In this way, the body weight and surface areas of an individual are kept internally consistent with across an individual's entire life.

In a similar fashion, the selection of records from separate studies is contingent on the values of inputs already assigned to the individual (see Figure 1). For example, the selection of a pesticide use record from the NHGPUS is made based on the type or home (single family or multiple family), setting (urban or rural), region, and presence of a yard. These inputs are also used to influence the selection of activity patterns for the individual. In this way, the data used for the characteristics of the residence, the pesticide use, and the individual's activities are taken from homes that have consistent characteristics. The selection of records or data for ephemeral inputs, such as dietary records and activity pattern records, are all drawn from data collected from individuals during similar seasons, in similar homes and in similar Census regions.

Correlation between inputs is also dealt with by modeling temporal trends for each of the inputs separately. For example, room sizes will remain constant until a person moves, while air concentrations will change from day-to-day based on pesticide use, season, and air-exchange rates.

Temporal correlations in source terms are managed by directly modeling the day-to-day changes in sources. For example, if a pesticide is used on one day, the following day's exposure is explicitly linked to the prior day's usage. Levels in air and on surfaces will be calculated in terms of the levels that occurred on the prior day. Typically, this is done by use of the preceding day's levels and a compound-specific decline rate. This type of linkage is not possible in models that glue together distributions of "single-day" estimates for different individuals (ILSI, 1998).

Finally, it should be noted that there are exceptions to this process of contingent modeling. In certain cases, characteristics adopted for a house may be inconsistent with subsequent values
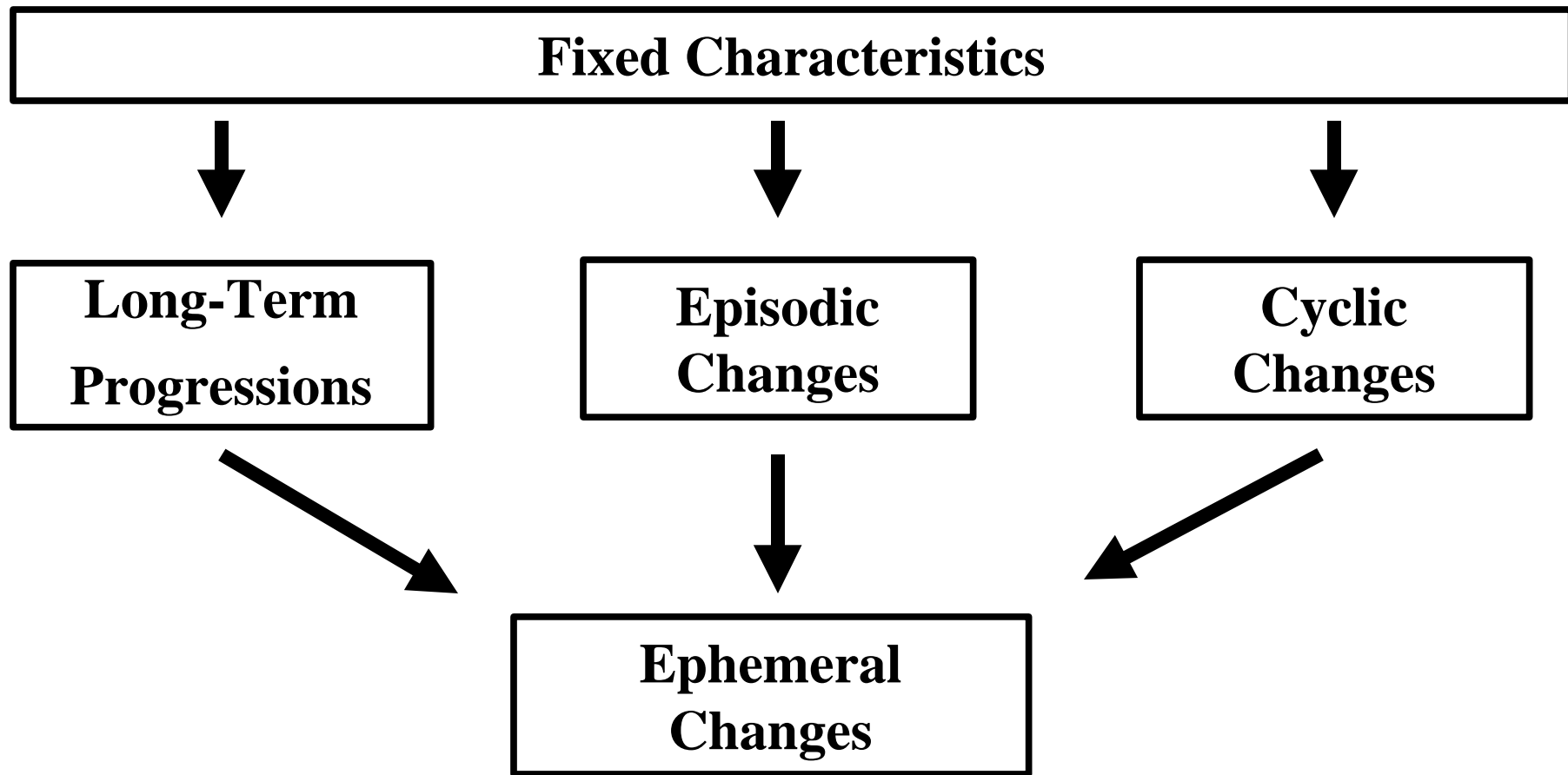
Fixed Characteristics

Long-Term Progressions	Episodic Changes	Cyclic Changes

Ephemeral Changes

Figure 1. Modeling Time Varying Changes in Inputs

associated with the selection of records of ephemeral inputs. For example, the information on tapwater source contained in the CSFII is not used in the model. Instead, the model uses the data in this input contained in the American Housing Survey. As a result, the model can assign a dietary record from an individual who had a private well to an individual in a home on a public water supply. This inconsistency was allowed since the source of tapwater in a home is unlikely to greatly influence the individual's dietary habits.

Modeling Specific Sources of Pesticide Exposure
Diet

Modeling exposure of pesticides that are presented in the diet follows the same concepts as presented above. The individual is defined and information about his or her dietary habits and the foods that are represented by those habits is applied appropriately. There is an abundance of information about dietary habits of persons in the United States, the food chain and pesticide use in that food chain. The biggest challenge for the model is to utilize the available data thoroughly and appropriately. The design of the utilized data must be clear to the system user so that any anomalies may be recognized and their consequences considered. Some of those will be pointed out in this presentation and are hoped to be the subject of other deliberations.

Assessment of exposure to pesticides in the diet must consider:
* How much food is eaten in a given period of time (by day, or even by hour or eating event), by whom (age, sex, socioeconomics, *etc.*), where, and in what form,
* Seasonal variations in the amount and form of the foods eaten, sources of those foods, and possible differences in pesticide use upon those foods, and
* Regional differences in the amount and form of the foods eaten, and possible differences in pesticide residues in those foods.

The inclusion of seasonal considerations into the assessment is critical. The individual may change her or his selection of foods, or the form of the food as eaten. Amounts of foods and beverages consumed may vary considerably with the season.

Food source is another major consideration. Many commodities in general distribution in the United States are available throughout the year, even though they are actually seasonal commodities. The source of the food changes with season. Citrus may be provided by California or Florida, depending on the month. Grapes may be provided domestically for certain months, but are primarily imported for other months.

Pesticide use patterns would naturally be different depending on the source of the food. Pesticide use in California will differ from that in Florida and may certainly be different from those in Chile or the Pacific Islands. The past practice of utilizing a single "percent crop treated" factor to account for these differences is not employed in this model. Rather, the present model applies the pesticide use profile to the foods eaten in a specific season. Where profiles of pesticide use are not available, the system user may make appropriate assumptions, including maximal use of the pesticide.

Food in the United States has become increasingly processed, both commercially and at home. Processing procedures can change the amount and nature of residue in the food. Additional

flexibility has been included in this model for using information on the effects of processing to modify the residue characteristics. A series of user-specified processing factors can be applied at the crop group, commodity, or food form level, as appropriate. This will be an increasingly important feature as we learn more about the effects of new processing procedures on the foods we eat.

Data Utilized

Consumption Data:

Information about the amount and form of the foods eaten is supplied by the United States Department of Agriculture (USDA) as surveys of consumption, conducted by the Agricultural Research Service (ARS). The earliest survey utilized by the EPA was a 1977-78 survey. The present model utilizes the 1989-91 Continuing Survey of Food Intakes by Individuals (CSFII). The 1994-96 survey, with its companion Supplemental Children's Survey, is to be incorporated in a future version of this software. The methodologies for those surveys have evolved significantly to improve accuracy and representativeness of the information.

The design of CSFII, in combination with the ARS Diet and Health Knowledge Survey (DHKS) component, is focused on assessment of the nutritional adequacy of the American diet and of the impact of individuals' attitudes and knowledge about healthy eating to their food choices and nutrient intakes. Thus, a wealth of information is presented about the individual, as well as the form and amount of food in her or his diet. This presents both opportunities and problems.

Information about each individual in CSFII supports linkage to the model's defined person. If we have, for example, defined an analysis for a given sex in a given geographical region, we can easily select the appropriate records from CSFII for inclusion in the analysis. Additional information accrued by ARS in these records, beyond that used to select a record, is preserved by the model for consideration at the end of the analysis. The system user can explore a given region of the exposure distribution to understand what "people" populated that section of the distribution and why their exposures may be what they are.

The user must understand, however, that certain statistical complexities arise from any use of these surveys for pesticide exposure assessment, as applied to EPA's policies. For example, the survey is statistically wholesome when used as a whole. Weights have been applied to account for underreporting within design cells, and the survey is very accurate at the mean of the distributions created for the whole population. Whenever population subsets are created, or the tails of the distributions are utilized, the statistical integrity is compromised, not reflecting any fault in the data or model using those data.

Food Recipe Data

The model uses a data set, called Recipe Files, that translates the foods, as eaten, to their ingredients and then to the raw agricultural commodities. These Recipe Files have traditionally been a private database, used commercially in proprietary exposure analysis systems. They have now been provided by TAS-ENVIRON for use in this model. The Recipe Files are visible, for the first time, in this model, allowing the accuracy of that database to be considered by the system user. The initial version of this model does not permit the system user to modify the recipe file, but subsequent versions may provide that opportunity.

The TAS Recipe Files are employed only for the 1989-91 CSFII database. The recipes have been customized to be appropriate for the food item descriptions in that survey. For the 1994-96 CSFII database, a new data set will be provided by USDA and EPA that translates the food items to the appropriate EPA raw agricultural commodities. That translation data set is presently under review and will be incorporated into a future version of the model with the 1994-96 CSFII data.

Pesticide Residue Data

The model permits the system user to employ a variety of databases included with the model, or provide her or his own estimates of the residues on or in raw agricultural commodities. The system provides a listing of the official EPA Tolerances on crops or crop groups. The latest organization of crops into groups (both formally recognized as rulemaking, and evolving toward rulemaking status) is provided in the model. This listing is also visible to the system user, and will be updated periodically in the subsequent versions. For those EPA-defined crops and crop groups, and the food forms of the crops, Tolerance definitions and values are available in the model.

The model also carries a second listing of official pesticide limits, the CODEX Maximum Residue Levels (MRLs). These values differ from Tolerances in two important ways. First, the value is derived from field residue trials using slightly different methodology. Second, the definition of raw agricultural commodities and groupings of those commodities are not aligned with the categories used by EPA. Crops are named differently and may be considered in different groupings or in different forms. A crosswalk from EPA RAC terms to the CODEX commodities and their MRLs is a feature of this model. This feature permits the user to explore the differences in the US regulatory standards versus the international standards.

The system user can also view the USDA's Pesticide Data Program's annual pesticide residue file. In the present version of the software, a direct link to the analysis program is not available, but the system user may view the PDP files, select residues to be applied to the analysis, and paste those selections into the analysis. Information about the structure and design of the PDP data set is provided to the user so that these values are used appropriately.

The system user can provide any other set of residue data for use in the analysis. The input structure provides an opportunity for data input at the crop group, crop, and food form levels. These inputs are automatically recorded, so that an accurate record of data and their use accompanies each analysis. In addition, the software has fields available for narrative notes by the system user to explain the logic underlying the selection of input data and analytical parameters.

Data can be in the form of individual values (deterministic data), as in the case of Tolerances and MRLs, or of distributions. Data distributions are used empirically by the analysis system, avoiding the inferences associated with parametric distributions. Future versions will accommodate parametric data distributions, but the analysis will explicitly identify them as such.

Residue Modifying Factors

Residues in foods can be changed by processing and the passage through the food chain from the field to the table. Those processes can modify the magnitude, as well as the nature, of the residue. Residues can concentrate during drying or concentrating processes. Conversely, they may be dramatically reduced during prolonged heating, bleaching or other common home or commercial food preparation processes.

The system user will be provided with a standard listing of concentration factors for those food forms that concentrate the food product by removal of water from the mass of the food. These factors are used as defaults—assuming the nature and total mass of the pesticide residue remained intact during the water removal from the food. These default factors are provided in the analysis system (and are visible to the system user) but may be modified by the system user if better information is available.

The system user will be presented with a format that easily accepts information about processing effects. These values can be applied to whole crop groups, to individual crops or to unique forms of foods. The name of the process that modified the residue is supplied in the format, or may be changed by the system user. Again, space for narrative is available to enhance the record of the analysis. Residue Modifying Factors are treated as deterministic values in the initial version.

When modifying factors and distributions of residues are employed, the application of "zeros" in the residue distribution will follow the Standard Operating Practices of the Office of Pesticide Programs of EPA. These operating practices are designed to avoid "double counting" of zeros in the pesticide residue distribution. Those practices are explained in detail in the software's accompanying documentation and in the EPA references.

Model Components for Dietary Exposure Assessment

The model has two distinct operating sections, the Food Residue File and the Exposure Assessment

The <u>Food Residue File</u> is a new feature, not employed in any other dietary exposure assessment model. It creates a unique distribution of residues for each food in the CSFII survey, as eaten. Each food residue file used in exposure analyses remains associated with the person's record and associated demographic data, awaiting "call-up" in the second operating section—the exposure analysis.

Developing an array of individual food residue files provides several advantages. First, before employing the array of data on crop residues, processing factors and other information, the system user is presented with a complete profile of the residues that seem to be associated with each FOOD as eaten by the CSFII respondent. The system user can examine this record to consider several key factors, and gain insight into the statistical issues that will be presented in the subsequent exposure assessment.

The residue values (deterministic or distributions) selected or provided by the system user are applied with the selected (or provided) processing factors to each food reported on each eating occasion in every CSFII record. The resulting distribution of food residues is tagged for identity

and saved in the record. The system user can see the food residue distribution and consider the characteristics of that distribution. For example, they can consider how broad the distribution may be (insight into variability), identify foods that seem to have extremely high residues and those where the distribution may be skewed.

This examination of the food residue files will provide insight into several issues:

- Unexpected high residue values from any food or types of foods can be spotted. The logic of the processing factor modifying factors may be reconsidered if it appears that they have been applied inappropriately.
- As the variability in the residue distribution increases, so too may the variability in individual food residue increase. The statistical consequences of these wide ranges of variability may be important to the system user. Typically, this situation would suggest that the Monte Carlo analysis in the subsequent exposure analysis should be run for many more iterations than with a situation where key foods are not displaying wide fields of distributions in the residues.
- Unusual distributions or skew of distributions will also suggest that a greater number of iterations may be needed in the exposure analysis.
- Potential "outliers" in consumption values or residue values can be identified at this point— by food. The genesis of these outliers can be considered by the system user prior to running exposure analyses. When the user is satisfied with the food residue file, the second operational section can be employed—the exposure assessment. The food residue file applied to that assessment is a "stand-alone" record, complete with an audit of the data elements contained therein and about to be employed in the exposure assessment.

The subsequent Exposure Analysis employs the Food Residue Files and the CSFII records for the amount of food eaten by the individual. Groups of records from the CSFII that are consistent with the description of the population described by the system user. These records define the dietary profiles of the individuals within that chosen population.

The CSFII records have been extensively examined to reveal patterns of dietary habits that may be characteristically similar within groups of individuals and different between groups of individuals. In this way, the CSFII data are used as they have been used traditionally—to provide actual data values for mass of food consumed by an individual. Now, however, the whole data base has been used to describe patterns of food consumption. Thus, we are providing a perspective on how a selected value fits into the overall pattern of consumption for that group. The patterns in consumption have been used to derive a sound empirical basis for matching the CSFII daily consumption records to the individuals in the overall aggregation of exposure. For example, we now know the age ranges in which eating patterns naturally fall. Shifts in these patterns occur with age, as the child selects a wider array of food items, goes to school or experiences periods of growth and intense activity. Those patterns are reflected in the model, rather than employing age groupings that were selected arbitrarily or to match previous traditions.

When the exposure analysis is run, a CSFII record is selected for each iteration from the appropriate pool of records (defined by the population selection and dietary pattern groups), to

supply the mass of each food item consumed. That value is applied to a value selected from the residue distribution for that food item (as was built in the Food Residue Files). This value is saved. The sum of such food item residues within any specified time period represents the mass of pesticide residue to which an individual is exposed by diet in that time period (a one-day time step is used in the initial version).

This process is repeated to create a series of exposure values across a period (season) for that individual. Thus, using the same individual's consumption record, for each food consumed, another residue value is chosen from each food residue file distribution. These new values are summed for all foods eaten. For each selected person this yields an array of exposure values that could have been encountered by that person across the season. The differences in these values represent the variation in residues in the foods presented in that season, which in turn reflects the sources of foods and use of pesticides at those sources. It also reflects the forms of the food eaten in that season; the processes employed in those food forms and other seasonal considerations.

The process is then repeated for another season, using the appropriate Food Residue Files for that season and a CSFII record for an appropriately defined person eating in that season.

This continues over the individual's lifespan. The completed individual exposure history represents one iteration of the assessment. The number of iterations that should be employed will be defined by the system user. That selection will be made considering factors such as:

- The motive for doing the assessment. If the user wishes to examine the temporal patterns in an individual's exposure history, she or he will want to examine the results of each iteration, and a only few iterations will likely be required. The general profile of a dietary exposure assessment becomes visible, but the statistical precision and representativeness for the population is poor. The assessment is speedy, however, permitting the system user to go on to another analysis, change the data or assumptions, and repeat the assessment.
- The statistical stability that is desired in the assessment. The variability that was seen in the Residue Data Files and the diversity of the modeled population, will affect the number of iterations required before the pattern of results for a population becomes stable. As the variability increases, the number of iterations that should be employed should increase to attain statistical stability in the exposure assessment.

There is no one "optimum" number of iterations that should be employed. The system user must consider what goes into the analysis and the statistical consequences of those inputs to determine how many runs to make.

Residential

The residential exposure assessment seeks to define the doses that each modeled individual receives from inhalation, dermal, and non-dietary ingestion on each day of his or her life. The dose rate equations used to derive these estimates are typically taken from the Residential SOPs. The goal of the initial version of the software is to avoid needless complexity, be as consistent with residential SOPs as is possible, and yet retain sufficient flexibility to take advantage of the information available on the modeled individuals.

As presented in Figure 2, the model is able to take data from multiple data sets and use it to define an individual's dose on a given day. The model is able to specify the following factors for each day of an individual's life:

- Personal characteristics (age, sex, weight, surface area)
- Behaviors (time spend in a micro environment, the specific activity, and breathing rates)
- Residence (number and nature of rooms, room sizes, air exchange rates, yard size, source of drinking water, and the presence of a pool, garden, and fruit or nut trees),
- Location (Region, urban/rural)
- Time (Season, week day or weekend)

As a result, the dose rate equations include many factors not currently considered in the residential SOPs.

As discussed above the inputs to the model are not fixed. The user is free to modify the inputs used by the model. The values of the inputs for the models are intended to capture the range of values that may be relevant across a population. This distribution of values for an input will generally reflect inter-individual variation in short-term or long-term measures of behavior. Guidance is provided for each input as to the nature of temporal or inter-individual variation that that should be incorporated.

The inputs are designed to allow the model to be used in multiple tiers of an exposure assessment, ranging from initial screening to complex, data rich assessments. In the initial screening assessments, the model can be run based on label information and preset assumptions for source terms. These assumptions can be replaced with actual data in higher tiered assessments. Even the initial screening assessment; however, will benefit from the detailed information on physiology, and exposure-related behaviors contained in the model. In addition, as the results of currently ongoing and planned research in exposure related behaviors become available, the model will be able to readily incorporate the data.

Tapwater

The software anticipates that the user will supply data on the occurrence and the distribution of concentrations in pesticides in tapwater. This assumption is parallel to the requirement in the dietary portion of the model for residue data. The model; however, allows the user to specify where contamination will occur and allow the use of a distribution of tapwater concentrations.

As with the residential exposure, the model defines many characteristics of each individual that are relevant to the evaluation of tapwater-related exposures. The location (Census region) and setting of the person's residence (urban/rural) is defined. The nature of the residence's source of drinking water is defined (private well, public water supply, bottled water or other.)

Because the model maintains this information, the user is able to specify water contamination on the basis of region, season, and type or water supply that is likely to be contaminated. For example, consider a pesticide that is known to contaminate surface water supplies during Spring run-off events and is used only in the western Census region. The model can limited the tapwater exposures to those individuals who live in the west and have tapwater supplies from surface
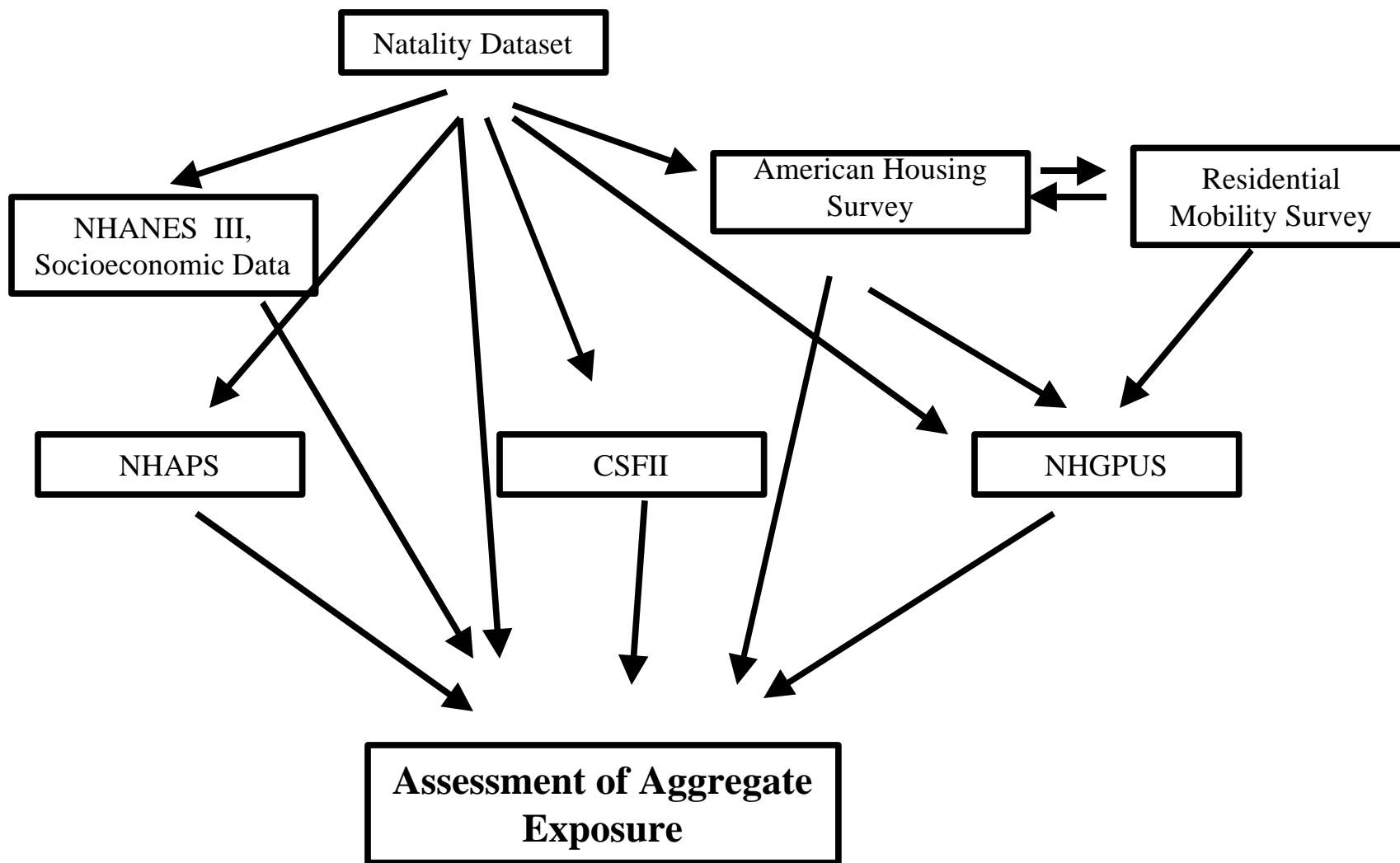
Figure 2. Databases Used in LifeLine™ Software

water sources. The exposures can also be limited to the Spring and can be modeled as a series of short-term periods of elevated exposures.

In contrast, a pesticide that is found to contaminate groundwater in areas near the application sites would be modeled differently. Only homes located in rural area and that rely on well water would be at risk of contamination. Since the location of the homes must be near the areas of application, it may be further possible to limit the fraction of the homes to those located in the counties where the pesticide is used. However, when a home was modeled as having the pesticide in the well water, the concentration would be modeled as having a relatively constant level over time.

The model will also allow the consideration of water treatment techniques (filtering, chlorination, ozonation, and use of activated carbon) on pesticide residues. To do so, the model will require estimates of the fraction of water treated by the specific method and the degree of reduction in concentration that is achieved.

As with earlier models, oral exposure to levels of pesticides in tapwater will be determined based on the amount of tapwater that is directly or indirectly consumed as reported in the CFSII surveys.

The model determines inhalation and dermal exposures during bathing/showering events as well as direct ingestion. These estimates are based on the Agency guidance on dermal exposure (Dang, 1996; EPA, 1992) and the McKone model of volatilization from showering (McKone, 1987; 1989).

Exposed and Non-exposed Individuals

By tracking the same individual over an entire life, the software will provide the user with the ability to better define and assess exposed and non-exposed populations. Many individuals are potentially exposed to pesticides during all or portions of their lives. However, on any given day individuals who are "potentially exposed" may or may not receive an actual exposure. For example, an individual may not eat any food item containing a residue, the residue may not actually occur in the specific item consumed, or the person may not be home when a pesticide is applied. However, on the following day these individuals can be exposed by these routes. In contrast, other individuals are truly unexposed because they live in regions or in homes where a pesticide is never used.

When the assessment is limited to a single random day in an individual's life it is difficult to differentiate between these two types of individuals. However, when an individual is tracked over time the two groups are readily differentiated. For example, if a pesticide is used in one region and the tapwater was the route of exposure, then the assessment could exclude individuals who never live in that region as truly unexposed. Individuals who have the pesticide in their water but on a given day do not drink any tapwater are included in the exposed population. This ability allows the user to evaluate individual's whose exposures are episodic while excluding individuals who truly have no potential for exposure.

Dose[7] Estimates

As discussed above, the model produces exposure histories for each modeled individual. The histories consist of the route-specific doses for each day of an individual's life. These histories provide the basis for deriving a number of different estimates of dose. These estimates include the following types of dose.

One Day Doses

Existing dietary and aggregate software produce distributions of daily doses that randomly occur on any day (that meets the criteria specified by the user) in different individuals. This distribution is duplicated by the model by pulling a single random day from an individual's entire exposure history. The day could be any day or can be restricted to those days that meet criteria specified by the user. For example, the user may specify that only days where the individual is between ages one and six will be sampled, or only days during Summer. A single day is pulled for each simulated individual. The distribution of the doses of these days taken from a large number of individuals is directly comparable to the results of the earlier software.

Doses over Multiple Days

In addition to duplicating the abilities of earlier software, the model also allows the evaluation of a group of consecutive days from an individual's exposure history. This group of daily records can be used to derive realistic estimates of average doses over any given period of time. For example, if a pesticide requires a dose to occur over 14 days in order to cause an effect, the model user can select automatically calculate the average dose over *each* consecutive 14-day period in the individual's life (*e.g.*, days 1 to 14, 2 to 15, 3 to 16, *etc.*). The distribution of these 14-day doses across individuals allows the user to accurately characterize the potential for risk. *The ability to create distributions of different dose durations will allow the tailoring of the exposure assessment to the needs of the toxicologist.*

Chronic and Lifetime Doses

This ability to select and calculate exposures over multiple days allows the derivation of chronic doses (one or more years in duration) for each individual. In addition, an individual's entire exposure history can be summed to produce a lifetime average daily dose. The distribution of these values across individuals allows for the first time insights on the distribution of the doses used for the assessment of chronic endpoints and carcinogenic risks. Thus, a single model run will produce distributions of dose for any length of time from one day to an entire lifetime.

Other Doses

The model also allows the tracking of the minimum, mean and maximum daily dose (or group of daily doses) for an individual. These values can also be tracked across individuals. For a population, of course, any percentile of exposure can be evaluated.

---

[7]   The software allows the analyst to include dermal and other absorption factors. Thus, both applied doses (often referred to as exposures) and absorbed dose (often referred to as doses) can be estimated using the software.

Modeling Doses to Children (and Other Age Groups of Interest)

Since the model calculates the dose that an individual receives on each day of his or her life, a single model run will produce estimates of dose for children, adults, and any other age group of interest to the user. The model allows the user to focus on doses for any age group of interest by specifying the specific ages that will be sampled for a random daily dose (or dose over a randomly selected set of continuous days). In addition, since the model tracks the same individuals as children and as adults it is possible to directly model the differences between children's and adult's doses in the exact same population.

Linkage to PBPK and PBPD Models

While not included in the initial version of the software, the route-specific doses in an individual's exposure history are ideally suited for use as inputs to physiologically based pharmacokinetic and pharmacodynamic (PBPK/PBPD) models. Once linked to such models, the exposure histories will produce histories of an individual's total body burden or concentration of a pesticide at the "target" organ, for each day of the individual's life. These estimates can be used to determine the potential of a pesticide to build up to unsafe levels in an individual's body. Such models could also be used to predict dose to the fetus and the nursing child.

When linked to pharmacodynamic models, the models will be able to track cumulative sub-clinical damage and recovery processes in an individual over his or her lifespan. This information, along with information on the individual (gender, current age and whether pregnant or nursing[8]) could allow the prediction of the actual potential for the occurrence of adverse effects in the individual's life.

Description of the Software

Model Design

The Software model is designed as a stand-alone computer program, written in C++. This has several key advantages in serving program goals:

- It runs far more efficiently than applications that require external programs such as Excel, @Risk, or Crystal Ball,
- It can be provided at lower total cost to the public, who need not purchase these applications,
- It is inherently modular (C++ is an object-oriented language) facilitating modification as the state of the science advances and as comments are received from the scientific and user community.

The model begins by selecting an individual's fixed characteristics, and then based on these characteristics, marches the individual through each day of his or her life. The route-specific doses received from diet, residential use of pesticides, and tapwater are calculated for each day and stored. Along with each day's doses, the model calculates rolling averages of dose for user-specified periods of time (2 days, 7 days, etc.) When modeling of one individual's entire life is

---

8   The current version of the software does not track pregnancy or nursing. However, data on the age specific probability of being pregnant and nursing a child are available and could be considered in future versions of the software).

complete, the software saves the measures of dose that are of interest to the user and moves to the next individual. This process is repeated for as many individuals as desired. At the end of the modeling of individuals, the program presents the distribution of the measures of dose across the entire modeled population to the user. In addition, data are automatically maintained on the details of exposure for every day in every individual's life where the exposure is unusually high. This facilitates the identification of factors that drive high exposures.

Two Modes of Operation

The model is operated in one of two modes. The first mode is an interactive use of the model. In this mode, the model quickly simulates the exposure history of a randomly selected individual within a defined population. This history will include a description of the daily intakes by route. The output of the simulation is displayed graphically. The output of additional individual's exposure histories generated from the same model inputs are produced and displayed at a single keystroke.

The purpose of this mode is to provide the user with insights on the temporal patterns of exposure, the relative importance of the different routes of exposure, and the ages where the highest levels of exposure occur. By reviewing a number of histories, the user is provided with insights on inter-individual variability. These insights can be used to determine whether the model inputs are correct and to provide a context for the estimates of the distribution of doses across individuals.

The second mode is the characterization of the distribution of total and route-specific doses received across multiple individuals defining a population. This mode is run as a "batch. " That is the parameters of model operation, including input data, are specified and results are returned after a number of individuals (specified directly or on the basis of some criterion of model completion) are simulated. Current software models operate in a similar mode.

Model Outputs

The output of the model occurs in several forms. In the interactive mode, individual exposure histories can be displayed. These estimates are also produced as file that can be saved for later analysis or exported for external analysis using Access, Excel, or other commercial software.

The population mode of operation produces an exposure report for each model run. This report can be printed out as a permanent record, with user selection of key report components. Report files maintain all vital data regarding how the estimates were generated.

This report contains the distributions of doses across individuals that the user had identified as being of interest. Thus, for example, the user could select distributions of average daily dose, maximum daily dose, maximum average dose over any 28-day interval, *etc.* Different distributions are developed for each of the averaging periods (from one day to a lifetime) as specified by the user. These distributions may be for the total dose (across all exposure routes) or dose from a specific route.

In addition, the report can include very detailed information on those that yielded the highest dose of interest-. This information includes:

- the individual's gender, ethnic background, and race
- the day and season the dose occurred,
- the age of the individual,
- the specific NHAPS and CSFII records used in generating the estimate
- dietary residues, tapwater level, and residential pesticide usage information, and
- all information regarding the residence at the time of exposure

This information allows the user to track the source and conditions that result in high estimates of dose, as well as to generate an automatic comparison to more traditional "random person-day" models.

Reports include a complete record of the data entered into the model and any user defined options. This information allows any other individual either to duplicate the inputs of the model and rerun the assessment, or to judge the credibility of the assumptions used in modeling exposure. Such information will be critical, not only for any regulatory submission, but also for any public presentation of model results.

Again, in addition to written reports, the model produces electronic files of the actual distributions of doses and the inputs. These files can be exported in formats that are not only accessible by the system, but can also be readily read by Access, Excel, or other commercial software.

Topics for Follow-up Presentations

The following is a list of possible topics for follow-up presentations to the SAP. The project team will be glad to

- Evaluation/Validation strategies
- Transition Rules
- Assessment of Children's exposure
- Variability, Uncertainty, and Sensitivity
- Predicting long-term from short-term measurements
- Linking pesticide uses in time
- Developing toxicological relevant measures of dose
- Linking to PBPK/PBPD models

References

Buck, R.J., K.A. Hammerstrom, and P.B. Ryan. 1995. Estimating long-term exposures from short-term measurements. J. Exposure Anal. Environ. Epidemiol. 5(3):359-373.

Dang, W. (1996) The swimmer exposure assessment model (SWIMODEL) and its use in estimating risks of chemical use in swimming pools. EPA internal guidance document.

Goodrum, P.E., G.L. Diamond, J.M. Hassett, and D.L. Johnson. 1996. Monte Carlo Modeling of Childhood Lead Exposure: Development for Use with the USEPA IEUBK Model for Lead in Children. Hum. And Eco. Risk Ass. (2) 4, pp. 681-708

Harrington, N.W., C.L. Curry, H.J. Carlson-Lynch, and P.S. Price. 1995. The Microexposure Event Modeling Approach to Probabilistic Exposure Assessment. Air and Waste Management Annual Meeting, San Antonio. Manuscript A443. May.

ILSI. 1998. Aggregate Exposure a Report to EPA/ ILSI Workshop, Ed. Olin, S, ILSI Press, Washington DC.

Keenan, R.E., M. Henning, P. Goodrum, M. Gray, R. Sherer, and P. Price. 1993. Using A microexposure Monte Carlo risk assessment for dioxin in Maine (USA) fish to evaluate the need for fish advisories. *Proceedings Dioxin '93 -- the Thirteenth International Symposium on Chlorinated Dioxins and Related Compounds,* Vienna, Austria.

McKone, T.E. 1987. Human Exposure to Volatile Organic Compounds in Household Tap Water: The Indoor Inhalation Pathway. *Environ. Sci. Technol*. 21:1194-1201.

McKone, T.E. 1989. Household Exposure Models. *Toxicology Letters*. 49:321-339.

McKone, T.E. and P.B. Ryan. 1989. Human exposures to chemicals through food chains: An uncertainty analysis. Environ. Sci. Technol. 23(9):1154-1163.

McKone, T.E. and J.I. Daniels. 1991. Estimating human exposure through multiple pathways from air, water, and soil. Reg. Toxicol. Pharmacol. 13(1):36-61.

Morgan, M.G. and M. Henrion. 1990. Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press. New York, New York. 332p

Muir, W.R., Young, J.S., Benes, C., Chaisson, C.F., Waylett, D.K., Hawley, M.E., Sandusky, C.B., Sert, Y., DeGraff, E., Price, P.S., Keenan, R.E., Rothrock, J.A., Bonnevie, N.L., McCrodden-Hamblin, J.I., 1998. A Case Study and Presentation of Relevant Issues on Aggregate Exposure, in Aggregate Exposure a Report to EPA/ ILSI Workshop, Ed. Olin, S, ILSI Press, Washington DC.

Price, P.S., C.L. Curry, P.E. Goodrum, M.N. Gray, J.I. McCrodden, N.W. Harrington, H. Carlson-Lynch, and R. E. Keenan. 1996. Monte Carlo modeling of time-dependent exposures using a Microexposure event approach. *Risk Anal.* 16(3): 339-348.

Price, P., J. Sample, and R. Strieter. 1992. Determination of less-than-lifetime exposures to point source emissions. *Risk Anal*. 12(3):367-382.

U.S. Environmental Protection Agency. 1992. Dermal Exposure Assessment: Principles and Applications. Office of Health and Environmental Assessment. Washington, DC. EPA/600/8-90/011F.