The LifeLine Group, Inc. is developing scientific software for broad public distribution on a non-profit basis. This document and the work products, ideas, concepts, data, calculations, graphics, code, logic and other intellectual creations contained therein are the exclusive property of The LifeLine Group, Inc.

Technical Manual for Version 2.0 of LifeLine™

What's New in Version 2.0

There are a number of changes, improvements, and extensions in Version 2.0. These improvements include: the ability to directly import existing dietary residue files, improved modeling of residential exposures, the ability to determine the aggregate and cumulative estimates of "Margins of Exposure" using the latest EPA guidance, the ability to investigate the exposures, doses, and risks on a randomly selected day, and improved graphics and tables. Version 2.0 achieves this performance without requiring users to purchase additional memory or high-end processors and without requiring users to wait hours or days for results.

LifeLine™ Version 2.0 is the version of LifeLine™ software that was used to assess cumulative risks from organophosphate pesticides for EPA. LifeLine™ 2.0 will allow all stakeholders to be able to duplicate and independently review cumulative risk assessments developed for regulatory decision making.

Version 2.0

Contents

CHAPTER 1. INTRODUCTION AND APPROACH

This *Technical Manual* is intended to provide a detailed description of how the LifeLine™ software operates. The document is intended to supplement the *Users Manual* and *Demonstration Case* included in the software documentation.

LifeLine™ is a simulation model that draws on data from a number of different nationwide surveys of exposure related factors and then uses a logical and consistent set of rules to determine which records are extracted and how they are matched to simulate an individual's exposure history. This document describes the process used to create the internal data sets and to develop the estimates of exposure, dose, and risk. The document also discusses alternative approaches that were considered in the development of the model. Finally, the document indicates how the current model could be extended in future versions of LifeLine™ to address other sources of exposure.

1.1 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 Environmental Protection Agency (EPA) exposure guidance such as *Risk Assessment Guidance for Superfund* (EPA, 1988) and the *Draft Residential SOPs* (EPA, 1999).

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] 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 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 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

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

(McKone and Ryan, 1989, McKone and Daniels, 1991). These models have allowed the incorporation of data from multiple surveys.

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, Wilson et al, 2000). The technique of Microexposure Event Analysis has been proposed for use in evaluating both aggregate and cumulative exposure (Muir *et al*, 1998) and was endorsed as a useful approach by the USEPA Science Advisory Panel, (SAP, 2000). This approach has formed the basis for all of the aggregate exposure programs, SHEDs™, CARES™, and TRIM™.

LifeLine™ Version 2.0 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. 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.

Modeling from birth does offer a number of benefits for the assessment of pesticide risks. The approach determines the risks to individuals who are born and spend their entire lives under a set of permitted uses for an active ingredient (AI). Thus, it allows the risk manager to determine whether an AI would pose unacceptable risks if the use of pesticide products containing the AI were to continue indefinitely. In addition, modeling lifetime exposures allows the assessor to compare doses received in a population of children and the doses received by the same simulated individuals when they are adults.

However, modeling individuals from birth is not a required feature of the LifeLine™ modeling approach. The model could have begun with cross-sectional data of the current population, and carried these individuals forward from their current starting ages and circumstances. Future versions of LifeLine™ may include the option of studying populations in this manner.

1.2 Modeling Time Varying Exposures using Transition Rules

LifeLine™ is a Monte Carlo (probabilistic) model of the aggregate exposure to pesticides that occurs to each member of a simulated 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.

LifeLine™ seeks to achieve this goal 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. The following is a general description of the approach with illustrations. A detailed description of how these rules were applied in LifeLine™ is given in the subsequent chapters.

1.2.1 CLASSIFYING MODELING INPUTS

Under the system of the transition rules used in LifeLine™, 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 are assigned 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.

1.2.2 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.

The major source of data on these fixed properties is the nation's birth records. The National Center for Health Statistics (NCHS) 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 U.S. 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. 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.

1.2.3 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 are based on the region and setting (urban or rural) of the mother's home, as well as maternal socioeconomic status (SES).

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.

1.2.3.1 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. 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 is correlated with the characteristics already assigned to the individual.

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.

1.2.3.2 Episodic Changes (Non-Periodic State Changes)[2]

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, etc.) then new values are adopted. The approach makes use of age-specific probability of episodic changes that are available from the U.S. Census.

---

[2] 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.

Mortality is modeled as the ultimate episodic change. The model determines 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.

1.2.3.3 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 birth date (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). Seasonal variation of residues in foods will also fall into this category.

1.2.3.4 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 in a food item consumed?

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.

1.3 Transition Rules

In LifeLine™, 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 are 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 may 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 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 allow 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.

1.4 Correlations (Associations) Between Inputs

The approach used in the design of LifeLine™ addresses 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 is 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). Records used in LifeLine™ include daily activity patterns records, dietary records; natality records, census records for mobility and housing, and records on residential use of pesticides. The using data for multiple inputs from 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 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. For example, the selection of a pesticide use record from the National Home and Garden Pesticide Use Survey (NHGPUS) is made based on the type of home (single family or multiple family), setting (urban or rural), and region. 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 and ages.

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 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.

CHAPTER 2. MODELING INDIVIDUAL CHARACTERISTICS

2.1 Introduction

As discussed above, LifeLine™ is a model that begins not with a description of the use of a pesticide or its residues, but with the characteristics of the individual exposed. In this chapter, we outline the process used to define those characteristics.

The characteristics that will be defined for each individual include:

- Permanent characteristics;
- Length of the individual's life;
- Physical characteristics including; and
    - ⇒ The height and weight of the person throughout their life;
    - ⇒ The total surface area of the person throughout their life;
    - ⇒ The surface of portions of the person's body (hands, arms, torso, etc.); and
    - ⇒ The "resting" breathing rates for the person throughout their lives.
- The characteristics of the home and family into which the person is born.

LifeLine™ Version 2.0 creates estimates of the inter-individual variation in exposures to AIs, and the resulting doses and risk in a population by constructing an internally consistent model of each individual in the population. This provides a credible basis for evaluating associations between individual characteristics, and minimizes the chance of generating "monsters" composed of implausible combinations of characteristics. This also allows LifeLine™ to meet EPA's goals of creating estimates that are demographically, geographically, and seasonally consistent (EPA, 1999).

This process begins by assigning the permanent characteristics of the individual. These values are assigned from a single birth record. Thus, the values will be internally consistent since they come from a record of an actual person.

The definition of the permanent characteristics is also the step where the user can define the nature of the population that will be evaluated in a model run. The user can use the model to evaluate persons with characteristics that are representative of the entire U.S. population, or limit the model based on sex, race, ethnicity, or socioeconomic status.

Once the permanent characteristics are determined, they are used to define other characteristics that change over time. The lifespan of the individual is simulated using age-specific mortality that is linked to the individual's race and sex. The height, weight, estimates of surface areas, and breathing rates are based on a model of age-related changes in these factors, taking into account, sex, race, and ethnicity.

2.2 Defining Individuals for Models of the General Population and Sub-Populations of Interest

LifeLine™ Version 2.0 begins the modeling of individuals from their birth. Accordingly, data taken from the NCHS database of birth records are used to determine the permanent characteristics of the individual.

2.2.1 USER DEFINITION OF THE MODELED POPULATION

The user defines a population to be modeled by restricting the permanent characteristics of the modeled individuals. Thus, the user may either choose a population to be representative of the U.S. population as a whole, or limit the analysis to a single sex, race, ethnic group, or category based on socioeconomic status (income quartile).[3] A description of this selection process is given in the *Users Manual* to LifeLine™.

[3] While socioeconomic status is not a fixed characteristic in the same way, as sex, race, and ethnicity, most members of the U.S. population tend to remain in a given income quartile for the duration of their lives. Future versions of LifeLine™ may allow the individual to transition to another quartile.

Once the user chooses the specific population that will be modeled, the system selects the primary permanent characteristics of the individual by a random selection of individuals from an appropriate subset[4] of the NCHS Natality (birth records) database for 1996.[5]

2.2.2 USE OF THE NCHS NATALITY DATABASE

LifeLine™ Version 2.0 uses the NCHS Natality database. The Natality data set contains an immense amount of information on the circumstances of the birth and the mother's obstetric history, and is representative of all live births in the U.S. population.

For the sake of efficiency, only the information that is used to establish an individual's permanent characteristics and the characteristics of the individual's initial residence has been extracted from each record.

- Sex
- Mother's race
- Mother's ethnicity
- Mother's education [used to infer Socioeconomic Status (SES)]
- Mother's age [used to determine initial residence]
- Census region of (mother's) residence [used to determine initial residence]
- Setting of mother's residence (rural / urban) [used to determine initial residence]
- Birth month

The NATALITY Database from NCHS

The Public-use Natality database represents all births occurring within the United States within a calendar year. Data for 1996 were published in July 1998 by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), US Dept. of Health and Human Services. Data were extracted describing 3,835,932 births. Each record contains complete birth records (not only data describing the child and its parents, but also relevant medical history).

[4] If the user places no restrictions upon the population, the entire dataset (more than 3.8 million records) is used.

2.3 Permanent Characteristics of the Individual

2.3.1 INITIAL DRAW OF A BIRTH RECORD

Based on the users definition of the population to be modeled, the Natality database is truncated to eliminate any records that are not of interest (e.g., if only females are to be evaluated, all records for males are excluded). Next, a record is drawn at random from the remaining set (by a random selection from the first array, followed by a linked selection from the second array), to identify five permanent characteristics of the individual (sex, race, ethnicity, birth date, and mother's education) and three additional key determinants of the individual's first residence (Census region, setting, and mother's age, see Chapter 3).

2.3.2 DERIVATION OF SES FROM MOTHER'S EDUCATION

Socioeconomic Status (SES) is a determinant that affects the type of residence and other exposure-related variables. As noted above, LifeLine™ Version 2.0 uses income quartile as the measure of SES, and makes the simplifying assumption that an individual's SES remains constant throughout life.

Ideally, the prediction of income for the child should be based upon data for both the mother and father, as well as considering family structure. Unfortunately, no income data are available from the Natality files, and education data are available only for the mother. While some of the other data (*e.g.*, mother's age) likely are relevant to family SES, we have not identified a method for incorporating this information into the prediction.

Accordingly, the assignment of the individual to a socioeconomic quartile is based on the information in the mother's education contained in the Natality database. The relationship of years of education to population income quartiles (*i.e.*, equal numbers of people in each quartile, nationally, not equal dollar ranges, such that half of the

[5] Version 2.0 does not allow the user the option of selecting the population based on region. The model allows individuals to move from region to region. Thus, the region is not a fixed characteristic of the individual and will vary over an individual's life.

population earns less than the 50th percentile income) has been derived from Grubb (1995), who analyzed Survey of Income and Program Participation (SIPP) data for 1984-1990. Grubb's tables document males and females separately, but do not directly address race or ethnicity. Grubb's Tables 1 and 2 have been used both to determine income quartiles for 1990 and to establish the relationship between education and income quartile.

While the rank ordering of education effects varies somewhat for males and females, and sex is a far stronger predictor of income than is education[6], the data for males and females can be categorized reasonably well into a single set of quartiles. Fundamentally, those who did not finish high school represent the lowest quartile, and those with only a high-school diploma the next higher quartile (males with a vocational certificate also fall into this category). The next higher quartile represents those with some college, but not a bachelor's degree, those with four years are just below the cutoff (men) or just above (women). The top quartile (the most diverse with respect to mean income as a function of education) is comprised of those with at least a four-year undergraduate degree.

Accordingly, the rules used in LifeLine™ to assign income quartiles based on the mother's education are as follows:

- Did not complete HS (years of education <12): Lowest income quartile
- HS diploma only (years=12): Second quartile
- Some college or post-secondary education (years=13-15): Third quartile
- Completed college or graduate school (Years >=16): Highest quartile

Once quartiles have been assigned at birth, they are treated as constant throughout the life of the individual (a simplification required by the absence of data on shifts in SES over a lifetime). Future versions of LifeLine™ may consider the probability of an individual

[6] At the 10th percentile, male income is about $7,000 higher than female (approximately double), the difference increases to $11,000 at the 50th percentile (slightly more than double), and to $13,000 at the 90th percentile (about 67% higher).

changing quartiles. Subsequent data on SES are analyzed in terms of the appropriate quartiles for use in each data set.

2.4 Lifespan / Mortality

The user has the option of either specifying a fixed lifespan for all individuals in the population, or using standard mortality statistics to predict an age of death for each individual. In the first case, LifeLine™ simply models the selected number of years of life for every individual. In the second case, mortality is modeled probabilistically for every individual on an annual basis (i.e., an individual's life is defined in terms of a whole number of years).

2.4.1 NCHS LIFE TABLES

The basis for modeling mortality is the Life Tables published by NCHS as part of the 1995 Vital Statistics of the U.S. These are the standard tables used to describe life expectancy at birth and at various ages. These tables categorize mortality as a function of age, distinguished by race and sex.[7] For a cohort of fixed size, the tables indicate the number of individuals expected to be surviving at the end of the first year of life, the number remaining after the second year, and so forth. The probability of an individual's dying in any year of life can be determined as the difference between the cohort size at the beginning of the year and at the end of the year, divided by the cohort size at the beginning of the year.

Because of the lack of robust data for individuals older than 85 years, LifeLine™ Version 2.0 is limited to the evaluation of exposures that occur in the first 85 years of individual's lives.

[7] Data do not appear to be published allowing one to estimate the effects of ethnicity or SES on mortality. The latter factor, at least, would *a priori* be expected to have a significant influence on mortality.

The Life Tables from Vital Statistics

The Life Tables comprise Section 6, Part A, Volume II of *Vital Statistics of the United States*, published by NCHS. 1995 Data were used as the most recent available at the time the databases were compiled. These tables fundamentally report the number of survivors in a defined cohort of 100,000 individuals at various ages from birth to age 85.

2.4.2 INHERENT PROBLEMS IN PREDICTING LIFE EXPECTANCY

All predictions of mortality (and its converse, life expectancy) face a common problem, they are based on individuals who have already died. The older the individual is at death, the less likely it is that the circumstances of his or her early life will be similar to those of a person who was recently born.

Thus, while statisticians and actuaries commonly use these data to predict life expectancy at birth, it is important to bear in mind that a person who died at age 85 in 1995 was born prior to the First World War and the concurrent influenza pandemic, and whose childhood predated the development of vaccines for diphtheria, pertussis, tetanus, polio, or smallpox. On the other hand, such a person had fewer opportunities to be exposed to emerging diseases such as AIDS, and or to early death in an automobile accident. One might well expect *a priori* that the actual probability that a baby born today will survive to age 85 is quite different than would be predicted from the life tables.[8] Notwithstanding, LifeLine™ follows standard practice in using these data to predict life expectancy. In addition, the model will also allow the user to disregard mortality and require that all individuals live to any user-specified age between 1 and 85.

[8] Obviously, the predictions of the tables have greater presumptive accuracy at older ages. The prediction of the likelihood that an 84-year-old will survive to age 85 is far less subject to historical demographic influences than the same prediction for a one-year-old.

2.4.2.1 A Problem of Mortality Data and Race

The Life Tables provide limited descriptions of mortality by race. White and Black are reported independently, but the category "Other" is reported only in combination with Black ("All Other"). While analyzing the data, we discovered that if one compared Black Mortality to mortality for All Other, non-black, non-whites were predicted to have *negative* mortality. In other words, a starting cohort of 100,000 individuals would *increase* in size (an obvious impossibility).

Discussions with domain experts at NCHS indicated that the discrepancy reflects a problem of data quality, which was apparently one reason that data for non-white, non-black mortality are not reported independently. As there was no way to correct this problem, the following rules were established for using the Life Tables to predict mortality in LifeLine™:

- Data for Whites and Blacks were used as reported
- Data for Other races were taken from "All Other" without adjustment, even though these data include a substantial majority of Blacks.

2.4.3 REGULAR PROGRESSIONS - PHYSICAL GROWTH

The description of an individual at any point in time is determined by permanent characteristics (like those described in the prior section), regular progressions, episodic states, cyclic conditions, and ephemeral events. The most obvious and important of regular progressions is the growth of each individual from birth through adulthood. A host of highly correlated factors that control exposure changes in a systematic fashion during this growth process.

2.4.4 DATA SOURCES FOR PHYSICAL CHARACTERISTICS

2.4.4.1 Height and Weight

NHANES III is the most complete published source of data about the actual personal characteristics of concern for the model, in that it:

- Reflects actual measurements under consistent conditions (as opposed to self-reported values);
- Includes data on a large number of individuals collected as a representative sample of the U.S. population; and
- Contains demographic data that allow the linkage of the permanent characterizations established for the individual from the Natality database, and that have been confirmed by in-person interviews with survey respondents (for other databases, these values may be either self-reported or inferred).[9]

NHANES III

This database, distributed by the National Center for Health Statistics (NCHS), contains *measured* physical parameters on a representative population of more than 31,000 individuals between the ages of 2 months and 90 years, collected between 1988 and 1994. No comparably large set of consistently measured data on physical characteristics of the population could be identified. The data support classification not only by age, but also by sex, race (white, black, other), and ethnicity (Mexican-American, Other Hispanic, Not Hispanic).

2.4.4.2 Body Mass Index, Skin Surface Areas

Body Mass Index is a well-established calculation using Height and Weight, and so can be immediately calculated for all ages once height and weight have been determined, without any additional data. LifeLine™ uses the formula provided in the NHANES III documentation:

[9] For example, CSFII imputes Hispanic origin based on surname.

$$BMI = W / H^2 \text{ (m)}$$

Where

W is weight (kg).

H is height in (cm).

Calculation of skin surface areas is far more complex, involving a series of regression equations that have been developed for a limited number of data sets. These regressions are discussed in some detail in Chapter 6 of Volume I of the revised US EPA's *Exposure Factors Handbook* (*EFH*, USEPA, 1997). The values for regression equations presented there are used for the prediction of skin surface areas for adults and children.

2.4.4.3 Inhalation Rates by Activity Levels

Most of the data have been collected and analyzed by Layton, as reported in the *EFH*. These data are the basis of a series of regression equations described below.

2.4.5 PATTERNS IN DATA ON HEIGHT AND WEIGHT

2.4.5.1 *A Priori* Knowledge Regarding Height and Weight

Establishing procedures for generating plausible estimates of the height and weight of an individual at any age must reflect not only available data, but also knowledge of basic physiological principles that govern the changes in height and weight as an individual ages. This is particularly important in light of the absence of accurate, accessible data on the height and weight progression of specific individuals and the cultural changes in patterns of obesity.

For example, it is widely accepted that height for an individual increases monotonically with age until adulthood, at which point it remains relatively constant. There may be decreases in older individuals, which presumably represent normal aging and pathological processes (e.g., osteoporosis). In available cross-sectional data (e.g., NHANES III), however, there are apparent decreases for older individuals. These may,

in fact, reflect differences in early-life conditions of different cohorts.[10] Such trends may well be exacerbated for populations of lower socioeconomic status, or in populations with large numbers of immigrants from countries with lower standards of living.

Similarly, it is well established that changes in an individual's weight over time (absent severe pathology), are likely to be constrained, and that it is difficult to predict weight based on other factors. Weight is, for example, clearly influenced by height, but the contribution of height to the prediction of weight is limited for adults (e.g., Burmaster and Murray, 1998). For children, in contrast, height and weight are fairly well correlated, reflecting the pronounced change in each as a function of age, as will be seen from the NHANES III data.

2.4.5.2 Limitations of Available Data

Although the NHANES III data appear to be the best available, they are cross-sectional in nature. This requires the use of inferential methods to describe the progression of height and weight of an identified individual. These methods are discussed below.

A second problem is that the data are limited in terms of the number of individuals in particular demographic subgroups (defined by sex, race, and ethnicity). These limitations reflect the attempt to capture the relative demographic composition of the U.S. population as a whole within a limited set of measurements, but have the result that the prediction of changes in height and weight is more uncertain for small demographic subgroups than for the majority groups. For example, NHANES III has data for 5,340 Male, White non-Hispanic individuals, but for only 256 Male, Other[11], non-Hispanic individuals.

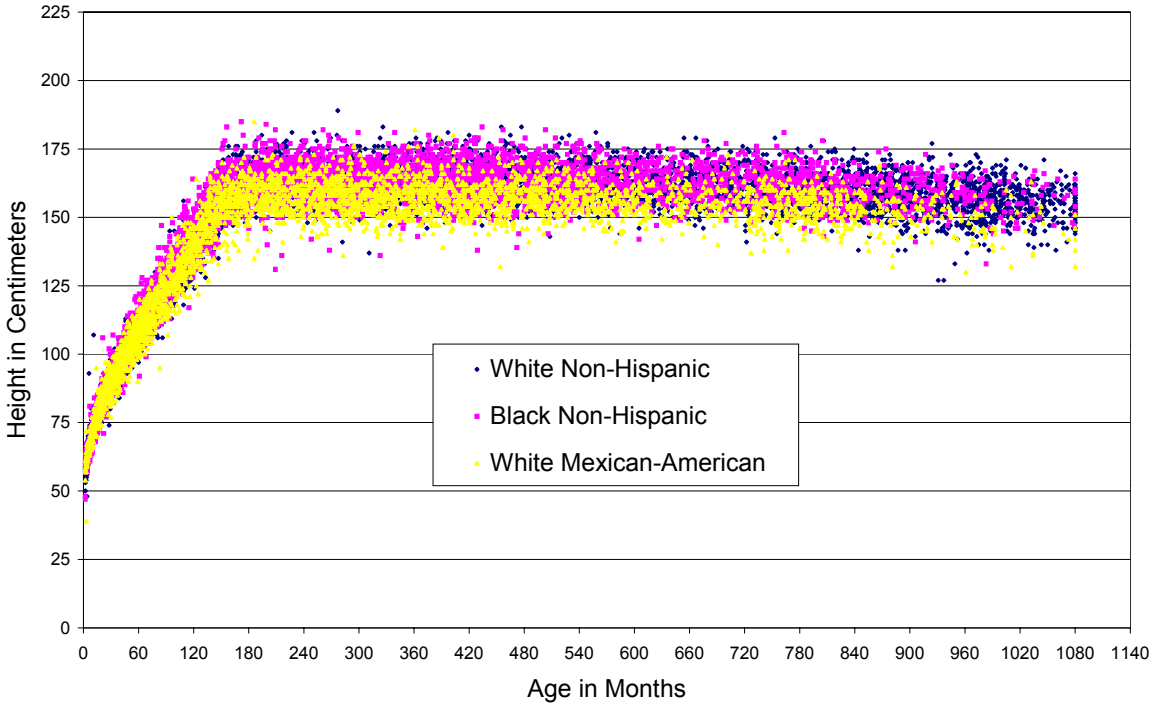
2.4.5.3 Data Characterization

In developing a procedure for predicting individual height and weight using the data contained in NHANES III, the data were first examined for trend, both for the entire population and for specific subgroups reflecting age, race, and ethnicity. The goal of this exercise was to accurately capture the NHANES III data, without compromising basic

[10] For example, poor nutrition experienced during childhood for those who experienced the Great Depression or were affected by the Second World War.

physiological principles (*i.e.*, height should not decrease except at advanced ages, weight fluctuations over time should be constrained to a reasonable rate).

Figure 2-1, Female Height as a Function of Age



As expected, both the central tendency and variability of height and weight increase as age increases, both for all groups and for each demographic subgroup (the data resemble a horn when height and weight are plotted as a function of age in three dimensions). Both the change in height and that in weight are clearly non-linear over a lifetime, primarily reflecting a transition from childhood to adult patterns.

Height shows a clear biphasic pattern, with a nearly linear[12] increase up to about 16 years (the actual inflection point varying among demographic groups, primarily reflecting

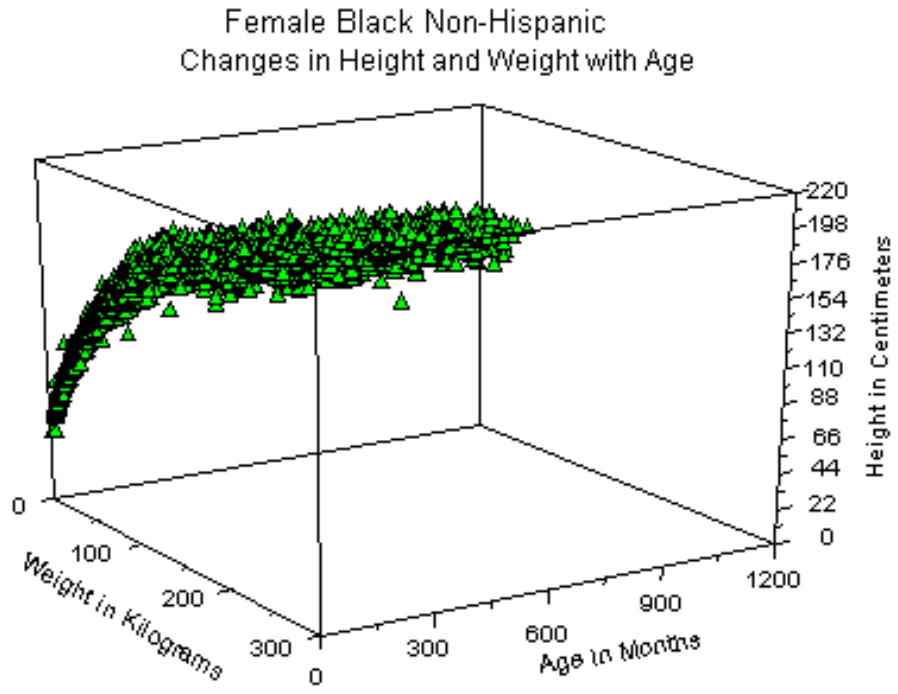
---

11 A category that includes racial groups other than White or Black.

12 Children's increase in height with age can actually be somewhat better modeled as two linear phases (with a breakpoint between 2 and 3 years), or by a non-linear function (at least one existing model of weight gain in children posits an initial hyperbolic phase, followed by a pattern better captured by a logistic regression), but there is a strong linear trend.

earlier attainment of adult height in females). Weight shows a similar pattern, except that a broader band of variability is evident. Height and weight are relatively highly correlated for heights less than four and a half feet, and very poorly correlated beyond that point. This clearly reflects different patterns of correlation in children *vs.* adults.

Figure 2-2



If one examines the data for children's height and weight[13], there is a clear trend for height to increase with age, with very constrained variability. The function is non-linear overall, but can be well approximated by two linear functions (an early steep slope and a later, more shallow slope, as noted above). In contrast, weight increases more gradually, with greater variability and with greater increase in variability as age increases. There are significant outliers in these distributions, notably at the upper end of the age range that may reflect adult patterns and the use of an arbitrary age criterion. Weight shows a

[13] Using an arbitrary threshold of 17 years = 204 months as the beginning of adulthood.

clear, non-linear correlation with height, which dissolves into variability at greater heights (again, this may reflect contamination with adult patterns).

For adults, height appears to be roughly constant, as expected, with a slight tendency to decrease at advanced age. Weight is extremely variable and very poorly predicted by age or height. For example, a linear regression indicates that for White Male Non-Hispanics, height accounts for only 6 percent of the variability in weight. For all heights with more than a few instances in the data set, weight ranges exceed 100 pounds. Age is even less useful as a predictor of adult weight, accounting for about one-tenth of one percent of the variation in weight for this "adult" data set.

2.4.6 PROCEDURE FOR DETERMINING HEIGHT

Initial consideration was given to employing a non-parametric approach, within groups defined by sex, race, ethnicity, and age, to assign heights. For example, for a Female Black non-Hispanic, one could assign the individual to a particular percentile for height (*e.g.*, 10^{th} percentile), and then determine the actual height value by selecting a height from the values within that percentile range (*e.g.*, between the 5^{th} and 15^{th} percentile) for that demographic group at any age.

It rapidly became apparent that cohort sizes in NHANES III data are too small to allow such a procedure, particularly for minority demographic groups. Moreover, in order to address both height and weight with this procedure, one would require even more data than were available for majority demographic groups. For example, to address weight classes *within* height classes (necessary to reflect the influence of height on weight) would require between 16 (for quartile classes) and 100 (for decile classes) categories for *each* demographic group at each age.

Because the non-parametric approach proved unworkable, LifeLine™ uses a parametric procedure, incorporating estimates of variability in the data to predict the height for an individual as a function of age. Subsequently, weight is modeled based upon the

individual's height and demographic subgroup. The development of the predictive model for height involves four phases:

- Segregation of data by demographic parameters (sex, race, ethnicity);
- Fitting models of central tendency for children and adults;
- Estimating a random factor to apply to the model for an individual that reflects population variability; and
- Adjusting the fit for the first and second years of life.

The errors that are introduced by this procedure primarily relate to the unsteady *rate* of growth of individuals over short periods of time. While the linear model offers a reasonable estimate of the population, and also (based upon clinical anecdote) of individuals over a span of several years, shorter periods of observation are very likely to be characterized by less stability in individuals (e.g., growth spurts and plateaus). These are well documented as a *normal* condition in the clinical literature (e.g., Porter et al, 1996). They are difficult to model accurately. In any event, the LifeLine™ model holds body size constant for any given year (as is typical in risk assessment), diminishing the relative contribution of this factor to errors in the analysis. None of these considerations would be expected to yield significant changes in any estimate of exposure or risk for either individuals or populations.

2.4.6.1 Fitting Linear Models of Central Tendency for Height

As discussed above, there are at least two distinct phases in the relationship between height and age. The approach taken here was to fit two linear functions, one for childhood and one for adults.[14] The intersection between these two functions is defined as the age (constant for a particular demographic subgroup) at which the individual is assumed to change from a childhood to an adult pattern (the transition age). By using a

[14] Because of the non-linearity in the children's data, a future improvement would be to either use a non-linear regression model for children's height, or to address three phases (early child, later child, and adult) with independent linear models. Based upon current data, this modification is expected to have a relatively small impact on the prediction of body weight, height, and surface area for any age group.

fixed transition age and a range of slopes relating height to age for children, a reasonable fit to the range of heights can be modeled for both children and adults. An additional small adjustment is used in the first two years of life, as regressions on children's heights consistently over-predict height in the first two years.

Height is modeled as the predicted value at the *midpoint* of each year. Thus, for the first year of life (age 0), the height at 6 months is used for the entire year. For the second year (age 1), the value at 18 months is used, for the third year, 30 months, and so forth.

Because the model addresses height and weight on an annual basis, the following rule is used to address the transition from childhood to adult patterns: When the evaluation age (in months) is less than the transition age, the actual age in months is used in the regression. When age first equals *or exceeds* the transition age (*i.e.*, on the first evaluation that occurs after the transition age, the *transition age* value is used, not the actual age.

For example, if the transition age were 153 months (12.75 years), the transition would occur between that for the 13^{th} year (age 12) and the 14^{th} year (age 13). For the 12-year-old, the value used would be 150 months. For the 13-year-old, the value used would be 153 months (the transition age), rather than 162 months. Thereafter, height is held constant, and adult patterns are assumed.

As might be expected, the linear model provided quite a good fit to the relationship between height and age for children, but a very poor fit to the relationship for adults (*i.e.*, once adult height is reached, age has very small, if any influence on height). Table 2-1 presents regression coefficients for height as a function of age in adults and children for the demographic subgroups in the analysis.

Table 2-1. Regression Coefficients for Height as a Function of Age				
Population	**Number of Children**	**Number of Adults**	**R^2 Child**	**R^2 Adult**
Female Black Hispanic	35	83	0.7985	0.0001
Female Black non-Hispanic	969	1860	0.8835	0.0009
Female Other Hispanic	50	113	0.8643	0.0057
Female Other non-Hispanic	54	87	0.8561	0.0002
Female White Hispanic	1071	1628	0.8773	0.0019
Female White non-Hispanic	831	1685	0.8903	0.0009
Male Black Hispanic	45	47	0.8886	0.0260
Male Black non-Hispanic	979	1388	0.8898	0.0003
Male Other Hispanic	75	122	0.8950	0.0016
Male Other non-Hispanic	54	75	0.8075	0.0512
Male White Hispanic	996	1501	0.8911	<0.0001
Male White non-Hispanic	808	1356	0.8853	0.0026

Visual inspection of the NHANES III data indicates that different demographic subgroups tend to reach adult heights at significantly different ages. These differences primarily are largest for sex, but with other factors play a role as well. In order to accommodate these differences, the linear regressions for adults were fit to slightly different age ranges for females and males so as to avoid using values in the age range where these transitions were occurring.

For both Females and Males, the childhood height regression function was fit to data for ages between 31 and 120 months (2.5 years to 10 years). For Females, the adult regression was fit to data for ages between 217 and 600 months (18 to 50 years), based upon visual inspection of the data. For Males, the adult regression was fit to data for ages between 241 and 624 months (20 to 52 years). Despite the selection of an adult age range to minimize any height loss associated with pathologies in elderly individuals, all of the adult regressions include a very small negative association of height with age (see above regarding potential reasons).

The linear regression of childhood height results in an overestimation of the heights for children less than 30 months of age. However, establishing a linear model on children

from 2 months[15] through 120 months of age, leads to an overestimation of heights in the area of intersection with the adult height function, and does not entirely eliminate the overestimation of heights for infants and toddlers. This reflects the non-linearity in height gain with age over this expanded age range.

In Version 2.0 of the model, this overestimation of children's heights is addressed by a small correction of predicted height for the first year (i.e., prediction age of 6 months) and the second year (prediction age of 18 months). This correction is discussed below. The use of three linear models or a non-linear regression could also have been used, however, the data suggests that the predicted values of height would not change to any great degree.

The following table presents the regression data for height as a function of age for each of the six modeled subgroups. Height is presented in centimeters, the slope is based upon age in months. Also presented is the predicted transition age in months for each group.

Table 2-2. Regression Data for Height as a Function of Age

Population	Child		Adult		Intersection (Transition Age) Months
	Intercept	Slope	Intercept	Slope	
Female Black Hispanic	79.70	0.52	160.11	0.00	155
Female Black non-Hispanic	74.96	0.58	164.34	0.00	154
Female Other Hispanic	75.21	0.52	157.64	0.00	156
Female Other non-Hispanic	76.44	0.50	158.25	0.00	164
Female White Hispanic	75.20	0.53	158.35	0.00	155
Female White non-Hispanic	75.94	0.53	164.61	0.00	166
Male Black Hispanic	73.90	0.62	176.16	0.00	164
Male Black non-Hispanic	77.43	0.55	176.40	0.00	181
Male Other Hispanic	75.60	0.55	170.86	0.00	171
Male Other non-Hispanic	77.96	0.48	178.19	-0.02	201
Male White Hispanic	76.60	0.53	169.88	0.00	176
Male White non-Hispanic	75.14	0.55	178.69	0.00	186

[15] The lowest age for which NHANES III data are available is two months.

2.4.6.2 Modeling Individual Differences in Height with Age

In order to ensure that the simulation of growth in height yielded a distribution of heights that reflected that seen in the appropriate (*i.e.*, demographically-matched) subpopulations of children and adults, the actual growth slope employed for each individual was selected from a normal distribution of slopes reflecting his or her subpopulation. The mean of this distribution was the slope of the child regression equation for the population, fitted as noted above. The standard deviation was selected to reflect the distribution of values around the mean slope, such that at the "transition age" where the two regression functions intersect, the distribution of growth slopes yielded a distribution of heights with the same standard deviation as that observed in the sample of adults used to calculate the regression equation. Table 2-3 presents for each group the height at the transition age predicted by the regression equations (and the gain in height from the origin of the regression), and the mean and standard deviation of height observed in the adult (217-600 or 241-624 months) populations used to determine the transition age.

Table 2-3. Height Data

Population	Transition			Adult Height Distribution	
	Age	Height	Gain	Mean	Standard Deviation
Female Black Hispanic	154.97	160.0	80.32	159.9	6.8395
Female Black non-Hispanic	153.89	164.1	89.10	163.6	6.3839
Female Other Hispanic	156.04	157.0	81.74	155.9	5.5829
Female Other non-Hispanic	163.53	158.1	81.68	157.9	6.7395
Female White Hispanic	154.80	158.0	82.77	157.4	6.0040
Female White non-Hispanic	165.77	164.3	88.37	163.9	6.3424
Male Black Hispanic	163.91	174.8	100.92	172.7	5.8819
Male Black non-Hispanic	181.74	176.6	99.19	176.9	7.0832
Male Other Hispanic	170.99	170.4	94.82	169.8	6.4805
Male Other non-Hispanic	201.34	174.8	96.83	171.3	8.4572
Male White Hispanic	175.96	169.9	93.26	169.8	6.5307
Male White non-Hispanic	185.58	178.1	102.96	177.3	6.6486

Because all of the variability in adult height for any subgroup is modeled as a function of the slope of the growth equation, the standard deviation seen for adults is applied to the difference between starting height (the intercept of the regression) and final height, rather than to final height *per se*, in order to estimate variability for the slope of the regression.

The following parameters are used to establish the population-specific distributions of growth slopes from which individual values are sampled.

Table 2-4 Central Estimate and Standard Deviation of Slope

Population	Mean Slope[16]	Standard Deviation
Female Black Hispanic	0.5183	0.0441
Female Black non-Hispanic	0.579	0.0415
Female Other Hispanic	0.5238	0.0358
Female Other non-Hispanic	0.4995	0.0412
Female White Hispanic	0.5347	0.0388
Female White non-Hispanic	0.5331	0.0383
Male Black Hispanic	0.6157	0.0359
Male Black non-Hispanic	0.5458	0.0390
Male Other Hispanic	0.5545	0.0379
Male Other non-Hispanic	0.4809	0.0420
Male White Hispanic	0.53	0.0371
Male White non-Hispanic	0.5548	0.0358

Once an individual has been assigned a particular slope for the (linear) function that predicts height from age, her height is then calculated according to that particular linear function up until she reaches the "transition" age, after which her height remains constant. That is, for any individual, height increases as a linear function of age (with a small correction for the first two years), but the initial height and the slope of that person's function reflects both her demographic subgroup and the variability seen in that subgroup.

2.4.6.3 Correction for the First Years of Life

As noted above, whether one uses all of the childhood data, or only those for children aged two-and-a-half and older, the regression of height on age systematically overestimates height for younger children. One way to address this would be to use three linear regressions in modeling height (very young children, children, and adults). A simpler approach has been employed here, reflecting the limited data on very young children in NHANES III.

[16] This is the slope from the least squares the regression equation.

The Table 2-5 presents the deviations of the regression predictions from observed values in the first two years of life. The predictions of the regression are compared to the (unweighted) mean height of all children in the corresponding year. Thus, for example, in the first year the predicted height at 6 months is compared to the mean height of all children between 2 and 12 months of age.

Table 2-5. Deviations of Regression Predictions from Observed Values for the First Two Years of Life

Population	Height (cm)			
	6 Months (2-12 mo)		18 months (13-24 mo)	
	Regression	Sample Mean	Regression	Sample Mean
Female Black Hispanic	82.8092	68.0	89.0288	82.6
Female Black non-Hispanic	78.4331	67.3734	85.3811	81.9439
Female Other Hispanic	78.3576	67.1	84.6432	81.0
Female Other non-Hispanic	79.4335	67.0	85.4275	77.9
Female White Hispanic	78.4065	67.9	84.8229	81.7
Female White non-Hispanic	79.1426	67.7161	85.5398	81.8116
Male Black Hispanic	77.5956	68.6	84.984	83.5
Male Black non-Hispanic	80.7008	68.7134	87.2504	82.2143
Male Other Hispanic	78.924	68.0	85.578	82.6
Male Other non-Hispanic	80.8497	68.1	86.6205	81.9
Male White Hispanic	79.7783	70.1	86.1383	82.7
Male White non-Hispanic	78.4696	69.7186	85.1272	83.1951

Based on the data above, the following correction factors were used to adjust the regression to predict height in the first and second years of life. In each case, the height predicted by the individual's regression (see below) is multiplied by the corresponding factor for her sex, race, and ethnicity:

Table 2-6. Correction Factors		
Population	**Correction to Regression**	
	6 Months	**12 Months**
Female Black Hispanic	0.8212	0.9278
Female Black non-Hispanic	0.8590	0.9597
Female Other Hispanic	0.8563	0.9570
Female Other non-Hispanic	0.8435	0.9119
Female White Hispanic	0.8660	0.9632
Female White non-Hispanic	0.8556	0.9564
Male Black Hispanic	0.8841	0.9825
Male Black non-Hispanic	0.8515	0.9423
Male Other Hispanic	0.8616	0.9652
Male Other non-Hispanic	0.8423	0.9455
Male White Hispanic	0.8787	0.9601
Male White non-Hispanic	0.8885	0.9773

2.4.7 PROCEDURE FOR DETERMINING WEIGHT

For each examined subpopulation, two observations were salient:

- For children, body weight is highly correlated with height, and less strongly correlated with age (for all but one small demographic subgroup, height accounts for more than 70 percent of the variation in weight); and
- For adults, there is essentially no correlation between body weight and age, and relatively low correlation with height. Indeed, the best R^2 value seen for any adult group (an anomalous value from a small population, more than twice as high as the next largest) indicates that less than half of the variation in weight can be accounted for by height.

This can be seen from the following table. Table 2-7 presents R^2 values for the regressions of weight against either age or height.

Table 2-7. R^2 for Regressions of Weight Versus Age and Height				
Population	**Child (31-120 Months)**		**Adult[17]**	
	From Age	**From Height**	**From Age**	**From Height**
Female Black Hispanic	0.269754	0.458518	0.045621	0.145668
Female Black non-Hispanic	0.622511	0.748232	0.044487	0.063024
Female Other Hispanic	0.540714	0.717328	0.077731	0.037246
Female Other non-Hispanic	0.584847	0.786687	0.035415	0.168132
Female White Hispanic	0.580776	0.741140	0.063032	0.079871
Female White non-Hispanic	0.607971	0.748632	0.037726	0.050126
Male Black Hispanic	0.635853	0.819370	0.000860	0.062946
Male Black non-Hispanic	0.640523	0.762111	0.004452	0.161819
Male Other Hispanic	0.662721	0.787064	0.041990	0.220602
Male Other non-Hispanic	0.633753	0.832634	0.005340	0.461174
Male White Hispanic	0.606058	0.744272	0.072793	0.181283
Male White non-Hispanic	0.664862	0.794893	0.025528	0.145272

Accordingly, different procedures are employed for modeling body weight in childhood and following the transition to adult stature.

2.4.7.1 Weight in Children

In accordance with the observed strength of the predictive relationships noted in the preceding section, a child's height (as derived above) is used as the primary predictor of her weight.

While there is a strong linear relationship between height and weight in children, particularly for those under ten years of age, this linear relationship tends to become significantly less reliable as height (and age) increases, reflecting the transition to the very poor relationship between height and age in adults. Moreover, the deviation appears to have a positive bias (*i.e.*, weights are generally higher than the predicted value), particularly once height exceeds 1.5 meters.

It has been noted in literature on adults (e.g., Burmaster and Murray, 1998) that the natural logarithm of weight is well fit by a normal distribution. And, in fact, a regression

relating the natural logarithm of weight to height in children not only offers a noticeable improvement in R^2, it also substantially eliminates the apparent bias in the deviations noted above.

Table 2-8. Correlation Coefficient for Weight and Log_e Weight Against Age and Height

Population	R^2 of Child Weight (kg)		R^2 of Log_e Child Weight (kg)	
	Age (Months)	Height (cm)	Age (Months)	Height (cm)
Female Black Hispanic	0.269754	0.458518	0.324346	0.558239
Female Black non-Hispanic	0.622511	0.748232	0.714310	0.843300
Female Other Hispanic	0.540714	0.717328	0.576686	0.750558
Female Other non-Hispanic	0.584847	0.786687	0.629451	0.840461
Female White Hispanic	0.580776	0.741140	0.655916	0.826170
Female White non-Hispanic	0.607971	0.748632	0.716842	0.861830
Male Black Hispanic	0.635853	0.819370	0.678853	0.845532
Male Black non-Hispanic	0.640523	0.762111	0.730778	0.863233
Male Other Hispanic	0.662721	0.787064	0.730751	0.861309
Male Other non-Hispanic	0.633753	0.832634	0.693564	0.874934
Male White Hispanic	0.606058	0.744272	0.692324	0.835353
Male White non-Hispanic	0.664862	0.794893	0.759443	0.892302

Accordingly, once the predictive equation for the child's height described above (linear function of age, with a normally distributed variability term for slope and correction for early years) has been applied, the corresponding central tendency of body weight is determined by applying the linear regression of the natural logarithm of weight on height (with a variability term as described below). Again, the linear regression and distribution are used to select a single value of the regression slope for each individual, which is then applied throughout the period of growth (i.e., each individual has a fixed slope assigned

[17] For females, 217-600 months, for males, 241-624 months.

for each regression). Table 2-9 shows the regression constants for predicting the natural logarithm of body weight (in kg) from height (in cm).

Table 2-9. Regression Constants for Predicting Body Weight from Height

Population	Child	
	Intercept	Slope
Female Black Hispanic	1.0098	0.0183
Female Black non-Hispanic	0.6771	0.0207
Female Other Hispanic	0.7772	0.0199
Female Other non-Hispanic	0.8197	0.0192
Female White Hispanic	0.6629	0.0211
Female White non-Hispanic	0.7112	0.0204
Male Black Hispanic	0.6507	0.0212
Male Black non-Hispanic	0.7937	0.0196
Male Other Hispanic	0.5149	0.0225
Male Other non-Hispanic	0.7909	0.0195
Male White Hispanic	0.7117	0.0207
Male White non-Hispanic	0.7340	0.0201

2.4.7.2 Variability in Weight at a Given Height for Children

While the above regression equations account for a sizable fraction of the variation in body weight, there is obviously significant individual variation that is not accounted for in the regression model. As would be expected from visual inspection of the transformed data, the absolute magnitude of deviations observed and the natural logarithm of body weight from the prediction of the linear regression increases with increasing age and height (i.e., there is more variability in the natural logarithm of body weight for older, taller children than for younger, shorter children). If expressed as a fraction of the increase in height (i.e., the value for the regression prediction corrected for the y-intercept), these deviations are roughly constant. In other words, the variability in the natural logarithm of body weight can be well modeled as variability in the slope of the regression on height.

Table 9-10 lists the characteristics of the distributions of relative deviations from which a slope adjustment multiplier for the regression of the natural logarithm of body weight from height for an individual is chosen for each demographic group.

Table 2-10. Distributions of Relative Deviations

Population	Mean Deviation	Standard Deviation
Female Black Hispanic	1.31E-05	0.090686
Female Black non-Hispanic	0.001785	0.057717
Female Other Hispanic	-0.01234	0.072263
Female Other non-Hispanic	-0.00012	0.050397
Female White Hispanic	0.002567	0.056242
Female White non-Hispanic	0.002298	0.052475
Male Black Hispanic	0.005379	0.057056
Male Black non-Hispanic	-0.00232	0.054043
Male Other Hispanic	0.00472	0.053953
Male Other non-Hispanic	0.005976	0.045738
Male White Hispanic	0.001924	0.056497
Male White non-Hispanic	-0.00091	0.047312

All of these distributions of deviations from the regression do however exhibit some degree of negative skew (a few large negative values and a larger number of small positive values) and positive kurtosis (the distribution is more tightly clustered around the mean than is the normal distribution). Because no suitable data transform or alternative distribution could be found for modeling individual variability, however, deviations from the regression of the natural logarithm of body weight on height were modeled as if they were normal.[18] This means that our modeled population will be slightly more variable than the population on which it is based, with an excess of small negative deviations and a deficiency of small positive deviations. The anticipated effects on predictions of body weight and height in a population will be extremely small.

As in the case of the prediction of height from weight, the prediction of the natural logarithm of body weight for an individual from height reflects both the basic regression and the random increase or decrease in the slope of that regression (a constant factor for

each individual) based on the deviation of observed heights and weights in persons in the same demographic category up to the transition age, as presented above. Table 2-11 presents the mean and standard deviation of the distribution of slopes for the regression equations of the natural logarithm of body weight on height.

Table 2-11. Mean and Standard Deviation of Slopes

Population	Mean Slope	Standard Deviation of Slope
Female Black Hispanic	0.0183	0.00166
Female Black non-Hispanic	0.0207	0.001195
Female Other Hispanic	0.0199	0.001438
Female Other non-Hispanic	0.0192	0.000968
Female White Hispanic	0.0211	0.001187
Female White non-Hispanic	0.0204	0.00107
Male Black Hispanic	0.0212	0.00121
Male Black non-Hispanic	0.0196	0.001059
Male Other Hispanic	0.0225	0.001214
Male Other non-Hispanic	0.0195	0.000892
Male White Hispanic	0.0207	0.001169
Male White non-Hispanic	0.0201	0.000951

2.4.7.3 Weight in Adults

The weight of an adult at the "transition age," like her height, is determined based on the regression equations applied to the individual as a child, and described above. Thereafter, one faces considerable difficulties in accurately predicting body weight.

A common perception, for example, is that adults generally tend to gain weight as they age. There is also some empirical support for this, although the available data suggest that the relationship of age and weight in adults is considerably more complex than the simple assumption stated above. For example, NCHS has recently released 1997 data on the prevalence of overweight (BMI >= 25) and obesity (BMI >= 30) (NCHS, 2000). These data indicate, based on self-reported height and weight, that the majority of the

[18] An alternative approach has recently been published, allowing one to use the S-distribution (with explicit terms for skew and kurtosis) to model variation (Voit & Schwacke, 2000). This may be appropriate for a future version.

U.S. population is overweight and one in five Americans is obese. Moreover, both men and women show an increased prevalence of both overweight and obesity in early and late middle age, relative to younger and older persons (data are presented in large age ranges).

These data are largely confirmed by examination of the NHANES data when organized in a comparable fashion. NHANES, however, indicates that the prevalence of measured overweight and obesity in women is noticeably greater than that in the self-reported data from NCHS presented above.

Table 2-12. Fraction Over Weight and Obese in NHANES and NCHS

Ages	Females				Males			
	Overweight		*Obese*		*Overweight*		*Obese*	
	1997	NHANES	1997	NHANES	1997	NHANES	1997	NHANES
18-24	33.3	38.9	12.1	17.3	41.5	36.9	13.9	11.6
25-44	42.9	57.3	18.2	30.3	63.7	59.6	19.2	20.5
45-64	55.0	70.9	24.6	38.2	70.7	68.2	23.0	25.7
>=65	50.5	61.4	18.0	25.4	59.8	60.6	14.4	18.2

These data must be contrasted with the observation, noted above, that age is a very poor predictor of weight when a linear model is applied to adult data ranging from age 17 to 50 in females or age 20 to 52 in males, a period over which the prevalence data on overweight and obesity suggest that there should be a consistent increase. Moreover, the slope of the function relating weight to age in each demographic group is quite modest.

In part, the contrast between these views of the data can be addressed by the fact that a substantial fraction of the population has a Body Mass Index (BMI) quite close to the classification criterion for overweight. Relatively small weight gains would accordingly shift significant numbers of individuals into the "overweight" category.

An examination of the patterns of BMI for each population subgroup is consistent with this interpretation. In this analysis, data for each subgroup were divided into five-year spans for ages 25-65, with additional groups addressing all persons from the transition

age for the group through age 24 and persons 65 or older. BMIs for each decile were plotted in each category. For the six subpopulations with substantial numbers of individuals (male and female of black non-Hispanic, white Hispanic, and white non-Hispanic), a consistent trend is seen:[19]

- There was a slight inverted-U relationship between age and BMI for most deciles;
- Much of this change in BMI with age reflected lower BMIs in persons under age 24 or age 65 and over;
- Median BMI was relatively close to the cutoff for overweight even for the youngest cohort, even for the lowest value (female white non-Hispanics), a 16 percent increase in weight would shift the median value over the cutoff; and
- For every demographic subgroup except female white Hispanics, more than 30 percent of the population was over the cutoff in the youngest cohort.

In light of the wide range of variation in BMI in each demographic group, the low power of age to predict weight or BMI in any group, the very shallow slope of the functions relating age to weight or BMI, and the absence of longitudinal data, the decision was made not to attempt to model weight change in adults. Attempting such a model was deemed to require an excessive reliance on conjecture.

As a consequence of this decision, there will be numerous small influences on many parts of the exposure analysis, relative to a model that did address such changes in BMI. Among the changes relative to a model that included consistent weight gain in adults, for example, would be the following:

- Dietary exposures will not show a consistent decrease with age in adults;
- Dermal exposures will not show a consistent increase with age in adults; and
- Inhalation exposures will not show a consistent increase with age in adults.

Obviously, with a more complex model of individual changes in BMI, more complex patterns of exposure change would be seen than with a model predicting consistent

[19] Data for the smaller subgroups are generally consistent, but a great deal more variable.

weight gain. Moreover, the magnitude of each of these changes reflects an interaction with age-based regressions of various exposure parameters described below. In any event, the available data suggest that the overall magnitude of all these changes would be rather small.

A more complex issue is that the model does not address an overall weight gain in the population over time, but only individual weight gain with aging. One plausible interpretation of the data described above is that the population in general is becoming more overweight, and a higher proportion of the population is obese, at all ages. This phenomenon is not addressed in Version 2.0 of LifeLine™.

Height and weight are recalculated for each individual on that individual's birthday. The assigned height and weight are those predicted for the midpoint of the coming year. Thus, at the first birthday, the weight expected at age 18 months is assigned, and is used until the individual reaches the second birthday.

2.4.8 CALCULATION OF BODY MASS INDEX

Once the height and weights are determined for the individual, the BMI is determined using the following equation (from the NHANES III documentation):

$$BMI = W / H^2$$

Where:

W is weight (kg)

H is height (cm)

The value of the BMI is recalculated for each year of the individual's life based on the individual's newly-assigned height and weight.

2.4.9 CALCULATION OF SKIN SURFACE AREAS

Different approaches for calculating skin surface area are used for children and adults, reflecting the different data sets and analyses available to support these predictions. For adults, the *Exposure Factors Handbook (EFH)* presents coefficients for regression equations that allow one to predict the size of specific body parts based on height and weight. For children, such regressions are not available. In part, this reflects a much smaller initial data set. For children, one is instead forced to predict overall body surface area based upon height and weight, and then to estimate body parts on the basis of the percentage of total body surface area that each part represents.

It should be noted that the data set upon which the prediction of body surface area from height and weight is small for children between the ages of 5 and 20 (the age range for which the regression in *EFH* is developed). The data sets used to estimate the percentage of total surface area represented by various body parts is miniscule.

2.4.9.1 Skin Surface Areas for Children

Whole – Body: Two separate regressions are used, as provided in Table 6A-1 of *EFH*. The first is applied to children under 5 years of age, the second to children between 5 years of age and the transition age determined for each group (see above). At the transition age and above, calculations for adults are used. The following regression equation is employed:

$$SA = a_0 * H^{a}_{1} * W^{a}_{2}$$

Where:

SA is the Surface Area of Body (m^2)

H is height (cm)

W is weight (kg)

a_0, a_1, a_2 are regression parameters for each body part (unitless)

The parameters are as follows:

Children under 5 years of age:

$a_0 =$ 0.02667, $a_1 =$ 0.38217, $a_2 =$ 0.53937

Children between 5 years of age and transition age for adulthood:

$a_0 =$ 0.03050, $a_1 =$ 0.35129, $a_2 =$ 0.54375

Body – Parts: The relative contribution of different body parts to a child's total surface area changes significantly over the course of development, notably with the decrease in the contribution represented by the head and the increase in that represented by the legs. While *EFH* tables data for every year of growth (Table 6-8 of the *EFH*), the number of children represented is so small that these predictions are not reliable. Instead, this model merges the data for children less than five years of age (a total of 14 children) and for children between the ages of five and eighteen (a total of 7 children), to yield two sets of body-part percentages. These are applied in the same manner as the regression equations for total body surface area reported above, yielding one set of predicted relationships for children under five, and another for children between the age of five and the transition age to adulthood.

Table 2-13. Percent of Total Surface Area for Various Body Parts

	Percent of Total Surface Area	
Body Part	**Ages Four and Under**	**Ages Five to Transition**
Head	15	10
Trunk	33	34
Arms	14	14
Hands	6	5
Legs	25	30
Feet	7	7
Total	100	100

This approach has the disadvantage of losing the fine temporal structure of the change in children's bodies from infant to adult form, but has the countervailing advantage of not attempting to model such changes precisely on the basis of one or two children, thus recognizing the significant inter-individual variability in children's bodies.

2.4.9.2 Skin Surface Areas for Adults

For adults, skin surface area for a series of body parts is calculated using the following equation, as presented in Table 6-1 in *EFH*[20]:

$$SA_p = a_{p0} * W^{ap1} * H^{ap2}$$

Where:

SA_p is the surface area of body part p (m^2)

W is the weight (kg)

H is height (cm)

a_{p0}, a_{p1}, a_{p2} are regression parameters for each body part (unitless)

EPA does not present parameters for all body parts for both sexes, and in some cases, no distinction is made between sexes. When a parameter is provided only for a single sex, or without designation by sex, that parameter is applied to both sexes in this model. Tables 2-14 and 2-15 present the regression parameters for males and females.

[20] This equation, and the parameters to fit different body parts, was derived in EPA (1985) *Development of Statistical Distributions of Standard Factors used in Exposure Assessments*, EPA 600/8-85-010 [PB85-242667].

Table 2-14. Regression Coefficients for Body Parts of Males			
Body Part / Body Area	**a_0**	**a_1**	**a_2**
Whole Body	0.02350	0.51456	0.42246
Head	0.0492	0.339	-0.0950
Trunk	0.0240	0.808	-0.0131
Upper Extremities	0.00329	0.466	0.524
Arms	0.00111	0.616	0.561
Upper Arm	8.70	0.741	-1.40
Forearm	0.326	0.858	-0.895
Hand	0.0257	0.573	-0.218
Lower Extremities	0.00286	0.458	0.696
Legs	0.00240	0.542	0.626
Thighs	0.00352	0.629	0.379
Lower Legs	0.000276	0.416	0.973
Feet	0.000618	0.372	0.725

Table 2-15. Regression Coefficients for Body Parts of Females			
Body Part / Body Area	**a_0**	**a_1**	**a_2**
Whole Body	0.02350	0.51456	0.42246
Head	0.0256	0.124	0.189
Trunk	0.188	0.647	-0.304
Upper Extremities	0.0288	0.341	0.175
Arms	0.00223	0.201	0.748
Upper Arm	8.70	0.741	-1.40
Forearm	0.326	0.858	-0.895
Hand	0.0131	0.412	0.0274
Lower Extremities	0.00286	0.458	0.696
Legs	0.00240	0.542	0.626
Thighs	0.00352	0.629	0.379
Lower Legs	0.000276	0.416	0.973
Feet	0.000618	0.372	0.725

2.4.10 ASSIGNMENT OF ACTIVITY-SPECIFIC INHALATION RATES

The final characteristic that is established is the individual's activity specific breathing rates. As discussed elsewhere in this manual, an individual's activities are defined in terms of the NHAPS activity records. Each of the specific activities (reading, sleeping, meal preparation, etc.) is defined by the user in terms of five categories of activity level (rest, sedentary, light activity, moderate activity, and heavy activity). The inhalation

rates corresponding to these five categories are established based on the individual's age, sex, and body weight.

A regression equation has been developed by Layton (1993) that allows the prediction of activity-specific inhalation rates based upon observed relationships between Basal Metabolic Rate (BMR) and Body Weight, and among BMR, activity level, and inhalation rate,). This equation has parameters that vary as a function of sex and age.

The equation is as follows:

$$IR = MET * BMR * H * VQ$$

Where:

IR is the inhalation rate (m^3 / day).

MET is the Metabolic Equivalent an Activity-specific BMR multiplier (unitless).

BMR is the basal metabolic rate in megajoules per day (MJ/d) [predicted by sex, age, and body weight].

H is the volume of oxygen consumed in production of energy (m^3/MJ).

VQ is the ventilatory equivalent, the ratio of minute volume to oxygen uptake (unitless).

The following values for MET are used:

Table 2-16. Metabolic Equivalents (METs) for Specific Activity Levels

Activity Level	MET
Rest	1.0
Sedentary	1.2
Light Activity	2.0
Moderate Activity	4.0
Heavy Activity	10.0

The value of H is 0.05 m^3/MJ and VQ is 27 as determined by Layton (1993).

BMR is determined (for a particular sex and range of ages) by the following equation:

$$BMR = (a * BW) + b$$

Where:

BW is Body Weight (kg)

a and b are constants that reflect sex and age categories.

Table 2-17 presents the values from a and b, as reported in *EFH.*

Table 2-17. Values of Coefficients a and b

Age Range	Males		Females	
	a	b	A	B
< 3	0.249	- 0.127	0.244	- 0.130
3 - 9	0.095	2.110	0.085	2.033
10 – 17	0.074	2.754	0.056	2.898
18 – 29	0.063	2.896	0.062	2.036
30 - 60	0.048	3.653	0.034	3.538
> 60	0.049	2.459	0.038	2.755

2.5 References

Grubb, W. N. (1995). *The Returns to Education and Training in the Sub-Baccalaureate Labor Market: Evidence from the Survey of Income and Program Participation 1984-1990*. MDS-765, National Center for Research in Vocational Education, Graduate School of Education, University of California at Berkeley

US EPA (1997). *Exposure Factors Handbook*. EPA/600/P-95/002Fa, August 1997, Update to *Exposure Factors Handbook,* EPA/600/8-89/043 - May 1989

Layton, D.W. (1993). *Metabolically Consistent Breathing Rates for use in Dose Assessments*. Health Physics, *64*(1): 23-36.

Burmaster, D.E. & D.R. Murray (1998), *A Trivariate Distribution for the Height, Weight, and Fat of Adult Men*, Risk Analysis, *18*(4), 385-389.

Porter, B., M. Kulikovsky, I. Shoham, & Z. Weizman (1996). *"Failure to Thrive" Over Diagnosis: Apparent Growth Deviation in a Healthy Infant Population*. International Child Health *7*(2), 7 pp.

Voit, E. & L.H. Schwacke (2000). *Random Number Generation from Right-Skewed, Symmetric, and Left-Skewed Distributions*. Risk Analysis, *20*(1), 59-71.

National Center for Health Statistics [NCHS] (2000). *Prevalence of Overweight and Obesity Among Adults in the United States*. National Health Interview Survey, Health E-Stats.

CHAPTER 3. MODELING RESIDENTIAL HISTORIES

Many of the factors that control exposure to AIs are closely tied to a person's residence. For example, potential exposure to AIs in residential settings will be strongly influenced by the following residential characteristics:

- Domestic water supply source;
- Presence or absence of a lawn or garden;
- Geographic variation in pest pressures (i.e., the likelihood that certain pests will require control in certain seasons);
- Household air exchange rates and seasonal variation in these exchange rates; and
- The probability of scheduled pesticide treatments that is independent of specific pest infestations.

Models that assume that these factors remain constant over an individual's life (the individual remains in a single home) introduce a potentially major source of error in exposure assessment. The size of this error will increase for estimates of exposures for periods longer than one year. Specifically, it will lead to over-estimates of long-term high-end exposures because situations that result in elevated exposures are treated as if they occur for longer periods than will occur for most individuals.

For example, a person who lives in a home with elevated AI concentrations in tapwater *may* reside there for an entire lifetime, but this is a rare event. Most individuals will move to another residence after only a few years. Failure to model mobility inflates the proportion of any modeled population that is presumed to have these continuing exposures, even if there is no change in the modeled central tendency of the population.[21]

[21] The opposite effect also occurs (inflation of the proportion of the population with abnormally low exposures), although this is generally of little concern in risk management decisions.

3.1 General Strategy – Mobility and Residential Characteristics

In LifeLine™, residential mobility for an individual is addressed on an annual basis. While this is a simplification of the actual patterns of mobility, there is a strong annual component in the timing of moves in real populations (reflecting standard lease terms, school calendars, etc.). In addition, the available data for mobility are collected on an annualized basis. This precludes any examination of the finer temporal structure of residential moves.

An individual's residential mobility is determined every year on the individual's birthday. This determination occurs in two stages:

1. Determining whether or not the individual moved, and the general nature of the move (e.g., from a rented urban multifamily unit to an owned urban single family unit) if a move occurs; and
2. Selecting a specific set of data (from an actual record) of the characteristics of the new home, by sampling from homes of similar type.

The evaluation of the frequency of each type of move is addressed by a set of frequency tables and the detailed characteristics of the home that a person moves to is then defined by a new set of housing records. These steps are fully explained in this Chapter.

3.2 Likelihood of Moving and General Description of a Move

3.2.1 TYPES OF RESIDENCE

Just as individuals can be described in terms of specific categories (sex, race, ethnicity, etc.), homes can be described in terms of a set of dimensions that are likely to influence either exposure opportunities, frequency of moving, or both. In LifeLine™, the basic description of a residence is addressed by four variables:

- Region (4 Census Regions are addressed);
- Setting (Urban or Rural, using MSA22 / non MSA as surrogates for these categories);
- Units (single or multifamily); and
- Tenure (owner or renter as occupant)

In addition, as in the case of descriptions of persons, these large categories (and the limited range of values in each) do not include all relevant factors. They appear, however, to be associated with significant differences in mobility, as well as exposure opportunities. Future research may suggest alternative approaches to categorizing residences that would offer advantages for exposure assessment.

3.2.2 DATA STRUCTURES

The basic data structure used to evaluate an individual's residential mobility is a frequency table (cut-point matrix) that gives the probability of either not moving or moving to a specific type of home based on the characteristics of the individual's current residence. There is one such table for each type of individual and each type of current residence, to reflect demographic differences in residential mobility.

Each of the mobility tables consist of 33 cells that describe potential for moving to different types of residence, (4 Census regions) by (2 settings) by (2 Units) by (2 tenures), the last cell describes the possibility of not moving.

Each cell contains a value $0 < x_i < 1.0$, with the sum of the cell's values ($\sum x_i$) constrained to 1.0. These represent the probabilities, extracted from the available data on residential mobility, that a person of a given type, living in a given housing type, either moved to any of the 32 housing types or did not move.

[22] Metropolitan Statistical Area

3.2.3 INPUT AND OUTPUT DATA

A specific module in the software determines the mobility of an individual on an annual basis. The module is activated when the age of the person changes, and results in a change (or no change) in the residence type of the individual. If the residence changes (even to a new residence of identical type), another module determines the characteristics of the new residence. If no move occurs, the characteristics of the residence remain unchanged.

Beyond describing the population to be modeled in terms of permanent personal characteristics (see Chapter 2), the user does not provide inputs to this function. For this module, input and output data have identical format, consisting of the description of the person and her/his residence type, in terms of the variables listed above.

3.2.4 STARTING CONDITIONS

The characteristics of the initial home of the modeled individual are selected based upon data on parents obtained from the birth (natality) records (See Chapter 2). These include:

1. Region (direct extraction from natality data)
2. Setting (urban or rural[23] from natality data)
3. Race (from natality data based on mother's race)
4. Age of mother (from natality data)
5. SES (quartile is selected based on mother's education level reported in natality data)

The process for this is described below.

3.2.5 DATA SOURCES

Personal characteristics have been extracted from the natality records, and are assigned to a modeled individual prior to modeling mobility. For the determination of initial residence, additional natality data concerning the mother and her residence are employed, as noted above. All other data on probability and nature of a move are extracted from either the Census *Current Population Survey(CPS), Annual Demographic Survey* (*CPS*)

files, or the *American Housing Survey* (*AHS*). These surveys are jointly performed by Census and the Department of Housing and Urban Development. *CPS* data for 1992 through 1999 were used.[24] Only data for persons who either did not move or moved within the U.S. are used, records involving migration into or out of the U.S. were discarded.[25] Data from *AHS* were taken from the 1993 survey results since this is the most recent year that includes information on previous tenure. *CPS* data provide the primary information on current residence and probability of moving. *AHS* data are used, in combination with *CPS* data, to describe the nature of a move. *AHS* data also provide detailed descriptions of individual residences.

3.2.6 FACTORS USED TO DETERMINE MOBILITY

3.2.6.1 Personal Factors

The personal characteristics that could be used to predict personal mobility (in addition to information on current residence - see below) are age, Income quartile (assigned as a fixed personal characteristic in this system), sex, race, and ethnicity.

A detailed graphical examination of the *CPS* database indicated that there are major differences in the extent to which each of these characteristics influences mobility. Income quartile had a major influence on the probability of having moved in the past year, as did age. In contrast, no consistent influence of sex, race, or ethnicity could be identified. Figures 3-1 and 3-2 present two views of the data on net lifetime mobility

---

23 Location inside or outside of an MSA is used as the surrogate for this variable.

24 Data for earlier years are not readily available electronically. Data for 1995 are excluded because the question regarding residential mobility in that year addressed the previous five years, rather than the previous year, and so the resulting data are not comparable to other years.

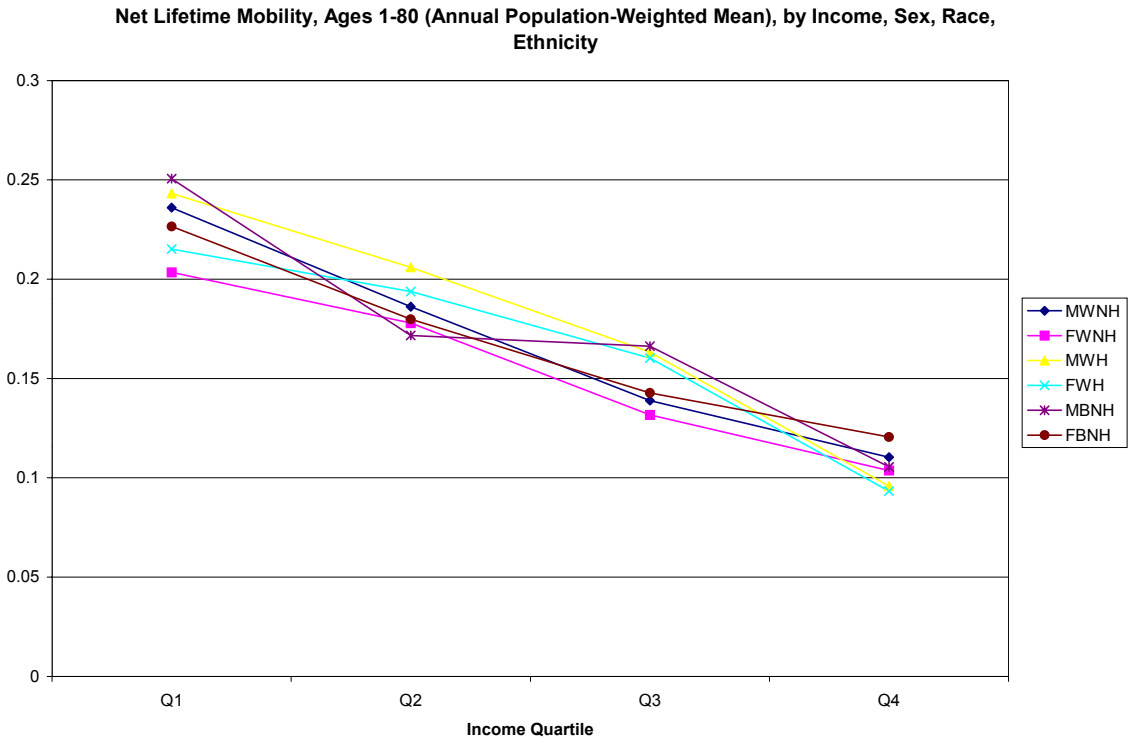
25 This simplification of the model will increase conservatism to some degree with regard to exposures occurring within the U.S. It may decrease overall conservatism to the extent that higher exposures are encountered outside the U.S.

26 Data for earlier years are not readily available electronically. Data for 1995 are excluded because the question regarding residential mobility in that year addressed the previous five years, rather than the previous year, and so the resulting data are not comparable to other years.

27 This simplification of the model will increase conservatism to some degree with regard to exposures occurring within the U.S. It may decrease overall conservatism to the extent that higher exposures are encountered outside the U.S.

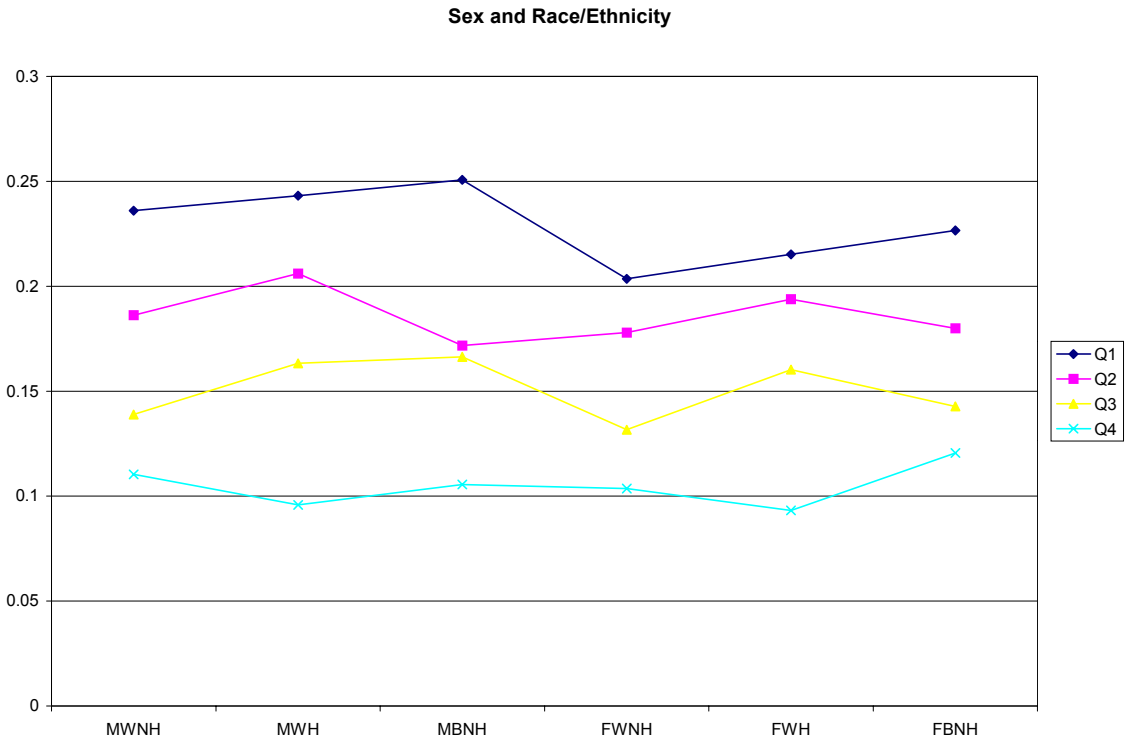
between the ages of 1 and 80, illustrating the effects of income quartile, sex, race, and Hispanic ethnicity.[28]

Figure 3-1



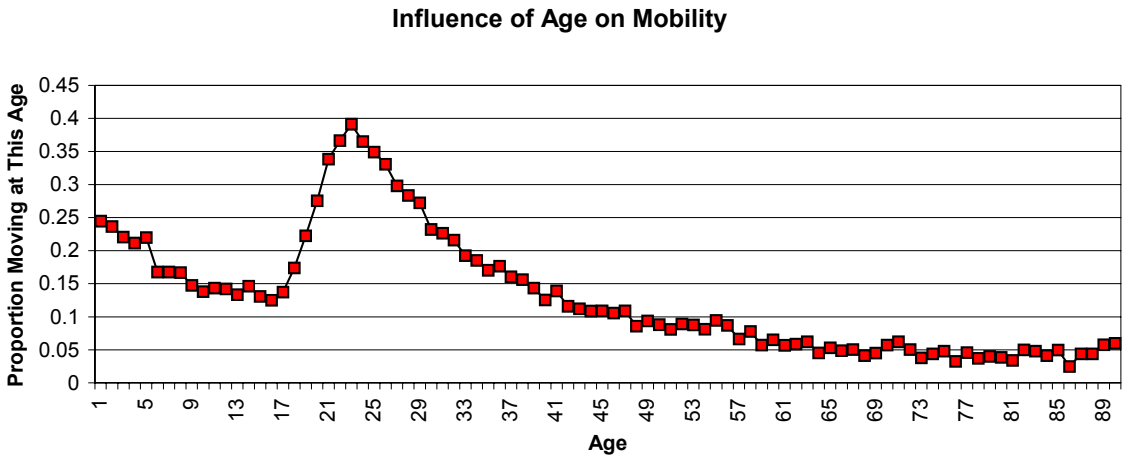
[28] This figure does not present black Hispanics, as the small sample sizes produced very variable results (but not consistent difference in pattern).

Figure 3-2



Age also has a major influence on the probability that an individual moved within the past year, as shown in Figure 3-3.

Figure 3-3



Because there are not enough data to support a year-by-year analysis of the influence of

age on mobility, and because the data suggest several distinct phases of residential mobility, age data were grouped into a series of six "bins," as described below. Even treating age in "bins," the factors listed above result in 288 different types of persons for whom a move has to be evaluated. Coupled with the description of possible starting and ending residences (see the following section), the number of possible moves to evaluate is very large, relative to available datasets.

3.2.6.2 Residence Types

CPS and *AHS*, when used together, support the analysis of residences (and, accordingly, of residential mobility) in terms of the four variables noted above, yielding 32 types of residence.[29] This set of 32 possible classes of residence provides for the possibility of 1,024 (32 X 32) different types of moves and 1,056 (33 X 33) moves if non-moves are considered.

3.2.6.3 Use of These Factors

The factors above (1,056 possible mobility outcomes for 288 classes of persons) result in more than 300,000 possible events to be modeled. Even with multiple years of data, none of the available data sets are rich enough to fully populate a table of probabilities for these events.

For example, the seven available years of CPS data represent approximately 765,000 records that are adequate to describe the *current* residence of a described individual *and* identify whether or not that individual moved in the past year. This is the largest data set addressing mobility. Even if these data were sufficient in kind to fully characterize a move (which they are not, see below), there are not enough records to describe the probability of all possible moves. Indeed, given the likelihood that different types of moves have very different probabilities of occurring, the data set would probably have no records whatever of several tens of thousands of the rarer types of moves.

[29] While, each database offers greater differentiation on some of these variables, the set of common variables for matching a residence are limited to these characteristics.

The strategy used to fill the tables with appropriate probability values is to make maximum use of available data for each category of cells. For example, there is a far larger dataset available for determining whether or not a person moved within the past year than there is for describing the nature of a move. Further, the data that describe moves are richer if one does not limit them to descriptions of moves that occurred in the preceding year.

Rather than attempt to make inferences based upon the small set of data defined by the intersection of two criteria (fully-described moves in the previous year), we determined probabilities for moving sequentially, using all the available data for each step. The procedure followed was:

1. To use all available records that describe *whether or not* a particular type of person in a particular type of residence was likely to have moved in the past year, in order to determine the relative probabilities of not moving and making any type of move,[30] and then;
2. To use all available records that describe the nature of a move (regardless of when it occurred) to determine the probability that a move of a given type of person from a particular starting point was to each of the 32 potential ending points.

Even with this approach, there are not enough data records to support the determination of an accurate estimate of the probability for each of the more than 300,000 types of moving events. Accordingly, each dataset was examined in turn, and the analysis was collapsed across those variables (personal and residential) that made the least contribution to variation in the probability of move events.

3.2.6.4 Other Data Issues

With the exception of the individual's age (or mother's age, in the case of initial residence), all data addressed in this module are categorical, both as regards the person (sex, race, ethnicity) and the residence (region, setting, number of units, tenure). Data on

[30] The implicit assumption being made is that the probability of moving is relatively constant over time for the general population and all subgroups.

age are grouped into age ranges (binned), reflecting patterns seen in overall mobility (i.e., whether the individual in the *CPS* reported a move in the preceding year). The data limitations noted above become insurmountable if one attempts to separately model the 85 different ages.

In the initial implementation of LifeLine™, the following age bins are employed, based upon inspection of the *CPS* data:

- 1-5 (an initial "peak" with mobility between 0.2 and 0.2531);
- 6-18 (the subsequent "trough" with mobility between 0.13 and 0.17);
- 19-32 (period when annual mobility exceeds 0.2), evaluated as two bins;
- 19-25 (period of accelerating and peak mobility32, mobility exceeds 0.2);
- 26-32 (period of decelerating mobility, mobility exceeds 0.2);
- 33-47 (mobility between 0.1 and 0.2); and
- 48 and above (mobility less than 0.1)

These bins reflect overall patterns seen in the data, although an examination of demographic subgroups (sex, race, ethnicity, and SES) indicates that there may be small inter-group differences in the apparent "natural" bin divisions. Thus, for example, not only do individuals in the highest SES quartile have lower overall probabilities of moving, the "spike" in mobility that appears to represent establishment of a new household occurs slightly later in life than it does for lower SES quartiles.

Another important factor to remember is that these data reflect mobility as measured in a particular set of residential types. These residences are defined to exclude a number of potentially important settings. For example, neither college dormitories nor Army barracks are included in these surveys, which could significantly distort the residential history of individuals between the ages of 19 and 25. Similarly, nursing homes, assisted-living facilities, and hospices are not included, so that the residential history of older individuals may not be adequately characterized.

---

[31] Probability that a move occurred in the year preceding the interview for *CPS*.

[32] Includes all years for which mobility exceeds one in three

For some combinations of race and ethnicity, there are too few records in the Census mobility files to define mobility in a robust fashion. This is particularly true for those ages where mobility is undergoing rapid changes (formation of new households). Some, but not all, of this data scarcity issue is addressed by binning on the basis of age and the use of all available years from *CPS*.

When there are insufficient CPS records to define the probability of the move for each cell of the probability matrix, the value from the next larger matrix, determined by the dominant characteristic, is used. In other words, the cell values are pooled and the pooled result is applied to each cell.

3.2.7 PROCEDURE FOR CALCULATING INITIAL RESIDENCE

This differs from calculating a subsequent move, in that natality data are the primary basis of determining residence (supplemented by CPS), and that house type is defined in two sequences. The same variables, of course, are used to classify the house type. The procedure consists of:

1. Extracting appropriate data from birth records concerning the mother's personal characteristics (age, SES, race, ethnicity) and residence; and
2. Inferring the two characteristics that are not present in the natality data (tenure, units) from data on women in residences who match this description

3.2.7.1 Housing Characteristics Taken from Data on Mother

As described in the section on permanent characteristics of the individual (Chapter 2), at the start of each modeled individual's life, data are drawn from the natality data (or the restricted subset chosen by the user). These are used to construct information on the initial residence, using the assumption that the child resides with the mother (at least for the first year of life):

- Sex, race, ethnicity, and region are determined in a random draw from the database.
- These factors are used to draw appropriate values for mother's age, setting of mother's residence (rural or urban), and mother's SES (derived from education), again from the natality data.

3.2.7.2 Determination of Tenure and Location

The natality data do not provide information on the mother's tenure (own or rent the home) or on whether the home is single-family or multifamily. These characteristics are defined using the following approach.

A table was constructed that describes the probability of the nature and type of home based on the individual's permanent characteristics. This table contains data extracted from the *Census CPS* data (current residence). It consists of 960 cells: (4 regions) by (2 settings) by (4 SES quartiles) by (5 age bins)[33] by (3 race bins) by (2 ethnicity bins).

In each cell, there are four probability values ranging between 0 and 1.0, and summing to 1.0. These represent the possible combinations of two values for units (single or multi-family) and two values for tenure (rent or own). The probability for each combination is determined by examining current housing data for women in the *Census CPS* data.

3.2.7.3 Using the Data

Based upon the description of the mother retrieved from the natality data, an appropriate cell in the birth housing data table is selected. A cut-point table procedure is used to establish number of units and tenure.

3.2.8 PROCEDURE FOR CALCULATING SUBSEQUENT RESIDENCE TYPE

After the assignment of the newborn to an initial residence, all subsequent moves utilize the same procedure. Each subsequent evaluation of a potential move[34] is based upon the individual's fixed characteristics, age, and current residence. Moves are evaluated annually, on the individual's assigned birthday.[35]

3.2.8.1 Determining the Move (Annual Basis)

The procedure for determining a move occurs in three steps:

---

33 It is assumed that no mothers are in the 1-5 year old range.

34 Including non-moves (remaining in the same residence).

35 This is clearly not a strictly realistic procedure (the distribution of birthdays across the year likely differs from that of residential moves (which appear to occur preferentially in warm-weather months). It is anticipated to have only tiny, nonsystematic effects on the estimation of exposures in any population.

1. Determine the appropriate mobility table and starting cell for an individual based on fixed personal characteristics (SES, sex, race, ethnicity), (new) age, and current residence type,
2. Generate a random number between 0 and 1.0.
3. Compare this to the cumulative probability (in consistent ordering) to determine which cell represents the person's new home (including the pseudo-cell of "no-move") [cut-point table procedure].

If the individual has made a move the model then goes on to select detailed housing characteristics based on the type of home the person moves into. Note that Version 2.0 of LifeLine™ does not address correlations in mobility in successive years. It is possible that such correlations could have an impact (Price et al, 1992).

3.2.9 PROCEDURE FOR POPULATING THE FREQUENCY TABLES

As noted above, in order to make maximum use of available data, the mobility frequency tables are populated in two steps. First, the probability of not moving in a given year (typically on the order of 0.75 to 0.80) is determined, using the richest data set that provides information on whether or not a move occurred in the previous year. Subsequently, for that fraction of the population that is determined to have moved, the probability of different types of moves is determined.

3.2.9.1 Identifying Whether or Not a Person Moved

The first step in populating the cut-point matrices for mobility is to extract data on the annual probability of moving, as opposed to remaining in the same residence. The database for doing so, as noted above, is richer than that for describing the nature of those moves that occur. This step is followed by partitioning the moves that do occur among the different possible types.

Data from the Census CPS database for 1992-1999 were used, excluding 1995 (which, as noted above, collected mobility data that were not comparable to those for other years). This database yielded a set of records describing whether an individual was living in the same or a different residence one year before the survey.

These records were sorted based on the individual's personal characteristics and characteristics of the current residence. The specific criteria were age bin, income quartile, sex, race, ethnicity, Census region, setting (rural or urban), type (single or multifamily), and tenure. Tenure is a description of the relationship between the individual (or parent if the individual is less than 18 years of age) to the residence (owner or renter).

As noted above, examination of the data indicated that neither sex nor race had any significant predictive power for whether or not a move had occurred in the previous year. No evidence of systematic effects was found, although for particular combinations of variables, results differed markedly from the overall pattern (i.e., for some combinations of age, income, and residential characteristics, males were different from females, or there were racial differences). In all examined cases, these variant points were found to represent a combination of variables with very few records. The limited number of records resulted in estimates that that would be significantly changed by the presence or absence of a few individuals. Similarly, the data for ethnicity were too sparse to indicate any systematic effects. Accordingly, sex, race, and ethnicity were not used in our analysis of these data, and all combinations of sex, race, and ethnicity were assigned the same values in the frequency[36]

All of the descriptors for housing type were retained. In general, Census region did not have a noticeable association with mobility in the previous year, with the exception that mobility in Region 1 (Northeast) was consistently lower than mobility in the other regions. In some cases, there were too few values to reliably assess the probability of move for a given combination of variables (i.e., fewer than 100 instances). In these cases, the overall value across regions was used to fill in the value for a particular region.

This analysis yielded 768 pairs of values (counts of the number of persons in each cell who did and did not move in the previous year). From these, 768 values for the annual

[36] This has the effect of reducing the 304,128 types of mobility events to be predicted to a set of 25,344 predictions, each of which is repeated for 12 demographic groups.

probability of not moving were derived. These values were replicated for all of the variables excluded in the analysis, to yield 9,216 estimates of the probability of not moving. These values directly filled the corresponding cells in each of the corresponding cut-point tables. They also constrained the sum of the values of the 32 other cells in each table.

3.2.9.2 Probabilities of Different Types of Moves

For any of the different classes of persons described by permanent characteristics, there are 1,024 moves (32 starting residence types and 32 ending residence types) that are theoretically possible, should a move occur. Of course, for any given move by an individual, there is a fixed current residence, so that only 32 moves are possible.

Data describing both current and previous residence are available from AHS for all descriptors except previous Census region, while CPS only has data on the location of the previous residence for movers (i.e., Census region and urban or rural setting). Accordingly, fundamental descriptions of the change (if any) in the type of residence during a move (i.e., with regard to rural or urban, single or multifamily, rented or owned) are obtained from AHS. These data describe an 8 by 8 matrix, representing counts for each combination of previous and current residence type.

3.2.9.3 Merging AHS and CPS Data on Mobility

AHS data are addressed to a representative set of residences, and do not reflect the distribution of the American population among different residence types. Thus the data on moves, while representative of the previous residence type for persons now residing in a particular type of residence, cannot be used directly to predict future residence type from current residence type. Rather, the AHS data must be adjusted to reflect the distribution of the population among residence types. The adjustment is as follows.

Table 3-1 below presents the format of a table that gives the possible moves between the different residence types. Each column represents the current residence type, while each row represents the previous residence type. For example, Column A and Row A could be envisioned as a rented urban multifamily home, while B represented a rented urban single

family home, etc. The shaded cells represent those moves for which the new residence is of the same type as the old residence (as opposed to the case in which the person does not move). Because each column in this table must sum to 1.0, the table has a grand total of 8.0.

Table 3-1. Probability of Mobility as a Function of Residence Type

		Current Residence Type							
		A	B	C	D	E	F	G	H
Previous Residence	A								
	B								
	C								
	D								
	E								
	F								
	G								
	H								

The data in AHS allow one to specify the proportions in each column, but not the relative magnitudes of different columns. That is, for a person in column A, AHS allows one to indicate the likelihood that the previous residence type was A, B, etc.

In order to predict moves (i.e., to utilize data across all columns for a particular row), it is necessary to have data on the relative magnitudes of the values in different columns. As noted above, these data cannot be obtained from AHS, but they are available from CPS for each population subgroup.

Accordingly, this initial table is replicated for all of the subgroups defined by the personal characteristics retained in the analysis and for Census region (6 age bins, 4 income quartiles, and 4 regions yield 96 subgroups). In each of these derivative tables,

the values in each column are adjusted using CPS data reflecting the distribution of the corresponding subgroup among housing types.

Thus, for example, the distribution of persons in the youngest age bin and the first income quartile among housing types is given in Table 3-2.

Table 3-2. Example Row

A	B	C	D	E	F	G	H
0.4	0.2	0.1	0.1	0.05	0.07	0.03	0.05

All proportions in Column A in the previous table would be multiplied by 0.4, while those in Column B would be multiplied by 0.2, and so forth. Each resulting table is thus constrained to a grand total of 1.0. More importantly, this transformation for each subgroup allows the tables to be used to predict the relative likelihood of 8 possible future moves (in terms of tenure, units, and setting), by selecting an initial row that corresponds to the *current residence*, and comparing the relative magnitude of the values in each column for that row.

As in the case of predicting whether or not a move occurs, this procedure assumes that past mobility histories are predictive of future mobility probabilities.

3.2.9.4 Adjusting for Absolute Probability of a Move

The tables above allow one to evaluate the relative likelihood of a move to the same or a new residence type (tenure, unit, or setting), but does not address the absolute likelihood of each type of move (as opposed to not moving). In order to make that determination, the values in each cell of a row must be adjusted so that the overall sum of the row is equal to the probability of moving for a person in the subgroup addressed by the table living in a residence of the type described by the row.

For example, the procedure described above might have yielded a row that summed to 0.09, for a starting residence of a rented urban multifamily home, in a table representing a person in the third age bin and second income quartile residing in Region 1. If the

probability of *not moving* in such a circumstance were 0.73, each cell value in that row of that table would have to be multiplied by 3[37] to yield a total probability of 0.27 moving one of the residence types described by that cell.

3.2.9.5 Intra and Interregional Moves

The procedure described above addresses all characteristics of a residential move with the exception of changes in Census region. In order to model these changes, CPS data on reported moves were analyzed to reflect the region of current and prior residence. CPS contains data on a starting and ending region for more than 450,000 moves (not all necessarily within the year prior to survey). These data indicate that the overall rate of moves between regions is relatively low. Of the 16 possible forms of intra- and inter-regional moves, intra-regional moves accounted for nearly 98 percent (97.98%) of all moves. Only three interregional moves accounted for more than one percent of the moves that originated in any region (only one accounted for more than 0.3 percent of moves overall).

Each of the tables described in the preceding step was replicated four times, to reflect the starting and ending regions of each move. The overall proportions in each table were assigned to reflect the probability of moving within or across regions (16 possible categories of move).

In light of the very skewed distribution of moves among the 16 categories overall, no attempt was made to analyze whether there were differences with regard to individual characteristics or residential characteristics other than region. The number of records where interregional moves occurred was too small to support a determination of the role of these factors.

3.2.9.6 Replication of Records for Characteristics Not Modeled

As noted above, the analysis of mobility data and resulting procedures for constructing the mobility tables reflected the lack of evidence for an influence of sex, race, or ethnicity. Records were replicated for all possible combinations of these variables, such

[37] That is the weighting factor has a value of (1.00 – 0.73) / 0.9 or 3.

that a male white Hispanic and a female black non-Hispanic (or any other pair) would have the same probabilities of moving to any particular new residence, assuming they were from the same age bin and income quartile, and the starting residence description was the same for each.

3.3 Assigning Additional Characteristics to an Individual's Residences

After the assignment of the initial residence, or the completion of any move, the characteristics given in Table 3-3. are defined by the model based on the results of the mobility module and the natality data.

Table 3.3. Known Characteristics of Residence and Individual Modeled

Residence	Resident
Census region	Age
Urban or rural setting	Income quartile
Single or multifamily	Sex
Owned or rented	Race
	Ethnicity

This section presents the process used to assign the remaining characteristics. The approach relies on the AHS database of residential characteristics. This database consists of records describing approximately 55,000 residences. Each record consists of a description of the location, setting, and type of home and income quartile of the owner/renter. The record also includes detailed information on the physical characteristics of the residence.

The model extracts from the AHS database all records that match the characteristics of the residence (except tenure) and the income quartile of the resident. One of these records is then randomly selected and used to characterize most of the remaining characteristics of the home. These records are then used with several additional databases to define the remaining characteristics of the residence.

3.3.1 HOME CHARACTERISTICS EXTRACTED FROM AHS

AHS records contain a large amount of information on residences, much of which is not relevant to the characterization of AI exposures. Therefore, the relevant data were extracted and entered into an internal database of the model. The data in the internal database include:

1) The nature of the foundation of the home;
   (Basement, Partial Basement, Crawlspace, Slab, Other)
2) The size of the lot;
3) The area of the finished space in the residence;
4) The number of the following types of rooms:
   a) Bedrooms;
   b) Bathrooms;
   c) Half-baths;
   d) Kitchens;
   e) Living rooms;
   f) Dens;
   g) Offices; and
   h) "Other" rooms.
5) The number of floors; and
6) Water supply of the home (public or private system, well, other).

As can be seen from this list, the AHS database does not include such relevant characteristics as the sizes of different rooms, the "footprint" of the home on its lot, or the presence or absence of a yard. Accordingly, a series of rules are applied to infer these more detailed descriptors from the data in the AHS record. The presence of gardens and certain plants is determined by data from a second survey and air-exchange rates from a third survey.

3.3.2 DETERMINING ROOM SIZES

The AHS record supplies the number and nature of rooms and the total finished floor space. This section presents a description of the process used to partition the finished

floor space among the residence's rooms. The first step in the process is the determination of the total number of "floor space units" (FSU) assigned to the home, reflecting the number and types of rooms that are listed in the AHS record. An FSU is defined as the measure of the relative size of each type of room to an arbitrary standard room. The finished floor space is assigned to each of the rooms in the home according to the number of FSUs allocated to that room. It is assumed that attics and basements are not finished space and are not included in the calculation.

The number of FSUs assigned to each type of room is based on a study of 21 single-family homes and 21 condominiums for sale in the Boston Massachusetts area in mid-1999. The size of the units varied from 700 to 5,900 ft^2 for detached homes and 440 to 3,900 ft^2 for the condominiums. The age of the homes ranged from "new" to more than 150 years old. The values were determined by assigning an arbitrary value of 1.0 to the primary bedroom of each residence. The remaining rooms are assigned FSU values using the following equation:

$$FSU_{ij} = RS_{ij}/RS_{Bj}$$

Where,

FSU_{ij} is the FSU for the i^{th} room in the j^{th} residence.

RS_{ij} is the room size of the i^{th} room in the j^{th} residence.

RS_{Bj} is the room size of the largest bedroom in the jth residence.

This process was repeated for all homes. The results of the survey generally showed consistent values for the FSU for each type of room. Data on sizes of bathrooms, halls, closets, stairs, and garages were not generally available in the home surveyed.

Table 3-4 gives the mean values of FSU for each of the various types of rooms.

Table 3-4. Floor Space Units for Different Rooms		
Room	**Floor Space Units**	
	Single-Family	Multifamily
Bedroom – Master	1.0	1.0
Bedrooms – Other	0.7	0.7
Kitchen	1.2	0.7
Dining Room	0.7	0.7
Living Room	1.3	1.1
Office, Den, Family Room, "Other"	1.2	0.85

These values were used to assign FSUs to each of the types of rooms listed in Table 3-4.

3.3.2.1 Apportioning Relative Room Sizes

The following rules are used in applying the floor space units.

3.3.2.1.1 Multiple Room Unit Values

For those rooms that have multiple possible allocations of floor units (i.e., bedrooms), the larger unit size is always applied to the first instance, and the smaller to all others. For example, the first bedroom is always a "master bedroom" and assigned a value of 1.0, while all other bedrooms are assigned a value of 0.7.

3.3.2.1.2 Special Rooms

Bathrooms and non-room (closets, halls, and stairs) space are allocated differently than are the principle rooms of the residence. The size of bathrooms is assumed to be largely determined by the standard sizes of plumbing fixtures and thus to be somewhat independent of the overall size of the home. In the case of non-room spaces, the model assumes that the total area occupied by these spaces is a fixed percentage of the floor space of a home.

3.3.2.1.3 Bathrooms

In single-family homes, if there is one bathroom, a size of 50 ft^2 is assigned to the room. If there is more than one bathroom, the first bathroom is assumed to be 100 ft^2. If the person modeled is an adult, he or she is assumed to use the larger bathroom, while a child is assumed to use the smaller bathroom. All additional bathrooms are assumed to be

equal to 50 ft^2. Half-baths are assumed to 12 ft^2. In multifamily homes, all bathrooms are assumed to be 50 ft^2 and half-baths are assumed to be 12 ft^2. Bathrooms are assumed to be defined by the plumbing fixtures and the space needed to access the fixtures. Certain bathrooms in larger homes may include Jacuzzis or designer fixtures. However, appropriate data on the corresponding room sizes have not been identified.

3.3.2.1.4 Hall/Stairs/Closets

These ancillary areas are assumed to be 15% of total finished floor space in a residence.

3.3.2.1.5 Room Usage

For those homes that have multiple possible allocations of floor units, adults are always assumed to use the larger room (master bedroom, master bath, etc.), while children use the smaller rooms.

3.3.2.2 Calculation of Room Sizes

The size of each room is calculated in the following manner. First, the areas of the Hall/Stairs/Closets, bathrooms, and half-baths are subtracted from the total residence area. Second, the FSU for each of the remaining rooms are totaled. The size of the FSU for each residence is calculated by dividing the remaining finished area of the residence by the total number of FSUs for the residence. Once the size of the FSU for the residence is calculated, then the size of each room can be calculated.

The net effect of this approach is to link the size of most of the rooms to the total size of the residential unit. That is, if the home is larger and the number of rooms constant, then the rooms are larger.

3.3.2.3 Example Case

Consider the case of an apartment with living room, dining room, kitchen, one bedroom, and one bath, with a finished area of 900 square feet. The following allocation of room sizes is made according to the rules presented above:

3.3.2.3.1 Halls/Stairs/Closets

0.15 * 900 = 135 square feet

3.3.2.3.2 Bathroom

50 square feet

3.3.2.3.3 Floor Space Units

Available footage = 900 – 135 – 50 = 715 square feet

Bedroom	=	1.0 FSU
Kitchen	=	0.7 FSU
Dining Room	=	0.7 FSU
Living Room	=	1.1 FSU
Total	**=**	**3.5** FSU

The number of square feet per unit is 715/3.5 or 204.29 square feet per FSU.

3.3.2.3.4 Floor Space in Square Feet (to nearest square foot)

The following sizes can be established for each room.

Bedroom	=	204 square feet
Kitchen	=	143 square feet
Dining Room	=	143 square feet
Living Room	=	225 square feet

3.3.2.4 Non-Finished Spaces

The AHS does not give any information on the size of unfinished spaces including attics, garages, and basements. There is insufficient data on the time spent and pesticide use in unfinished attics to allow the model to evaluate exposures in this space. Accordingly, the model does not attempt to characterize the sizes of attics.

The sizes of basements and garages are defined in the following way.

3.3.2.4.1 Basements

The AHS provides information on total finished space and the number of floors in a home. Floors are defined as levels above the basement. Using these data, the model

determines the footprint of the residence by dividing the area of the finished space by the number of floors. Basements are assumed to have an area equal to the footprint of the home and half basements to have an area equal one half of the residence footprint. The following equations are used to estimate the size of the basements and half basements,

Basement = Finished Space/Floors

Half Basement = Finished Space/(2*Floors)

Where

Basement is the area of the basement (ft^2).

Half Basement is the area of the half basement (ft^2).

Finished Space is the area of the finished space (ft^2)(taken from the AHS).

Floors is the number of separate floors in the home (taken from the AHS).

3.3.2.4.2 Garages

The model assumes that the purpose of a garage is to hold one or two cars. The available data do not allow the differentiation between one- and two-car garages. Therefore, we have assumed that garages are equal to a small two-car garage. A garage is assumed to be 25 ft by 25 ft.

3.3.2.5 Room Dimensions

Once the model has determined the floor area for each room of a residence, these data are used to determine the length of the room's perimeter and the room's volume.

The relationship between the room perimeter and room size is not constant. In reality, the shape of a room's perimeter may range for simple squares, to complex shapes with both straight and curved walls. The model attempts to deal with this variation by assuming that the relationship between perimeters and floor can be modeled by assuming that rooms are rectangles and allowing the user to specify a ratio of the length to the width for the rooms. By increasing the ratio, the length of the perimeter for a given floor area is increased.

The default assumptions are that all rooms have a length:width ratio of 2:3. This assumption is based on the observation that most rooms are not square. Hallways are assumed to have a length:width ratio of 10:1. Both values may be modified by the user.

The perimeter of the room is determined from the room size and ratio of length to width, a:b:

$$P = 2 * (a+b) * A / (a*b))^{0.5}$$

Where,

P is the perimeter of a room (ft).

A is the floor area of the room (ft^2).

a and b are the measures of the length and width of the room.

Since many rooms have more complex shapes than simple rectangles, the ratio of a to b might be increased to account for the higher perimeters.

The heights of all rooms are assumed to be 8 ft. Thus, the volume of the room is defined as follows:

$$RV = A* H$$

Where,

RV is the room volume (ft^3).

A is the floor area of the room (ft).

H is the room height, which is assumed to be 8 ft.

3.3.2.6 Room Dimensions - Units of Measurement

The basic data from the American Housing Survey on dwelling size are in square feet. These units are presented in the "View Analysis Results" display, and are also used in the LIVES.DBF log file for total house area.

In calculating exposures, however, room areas and volumes must be converted to square meters or cubic meters. The log file RESIDENC.DBF contains values in these units. Thus, for example, the same house will be represented in total area in LIVES.DBF as 1000 {ft^2}, while the constituent rooms (microenvironments) in RESIDENC.DBF will sum to an area of 93.5 {m^2}.

3.3.3 PRESENCE OR ABSENCE OF A YARD, GARDEN, FRUIT TREES

The presence of certain plants in the home are assigned based on data from National Home and Garden Pesticide use Survey (NNHGPUS). When a record is pulled for a home that record determines whether the home one or more of the following fruit trees, nut trees, or grape arbor, or whether the home has a vegetable garden.

1. If the unit is not a single-family dwelling, it is assumed an apartment without a yard[38].
2. If the residence is a single-family dwelling (either attached or detached) then the area of the ”footprint” (area of the foundation) of the residence is calculated. As discussed above, the footprint is defined as the size of the finished space of a residence divided by the number of floors in the residence. If the lot size minus the foundation area is less than 100 ft^2 then the program assumes no yard. If the lot size minus the foundation area is greater than 100 ft^2 then a yard is assumed to exist. The basis for this limiting assumption is that some portion of a lot is taken up by fencing, sidewalks, driveways, and gutters.
3. If neither of the above are true, the size of the yard is calculated as:

$$Y\ (Turf) = LS - (FS/F) - 100$$

[38] Recreational areas may exist for such dwelling but the apartment owners not the renter will manage the use of pesticides in these areas. As discussed in Chapter 4 of this manual Version 2.0 of LifeLine™ does not include data related to the use of pesticides in these areas.

Where,

Y is the size of the area on the lot covered by grass.

LS is the lot size (taken from the AHS).

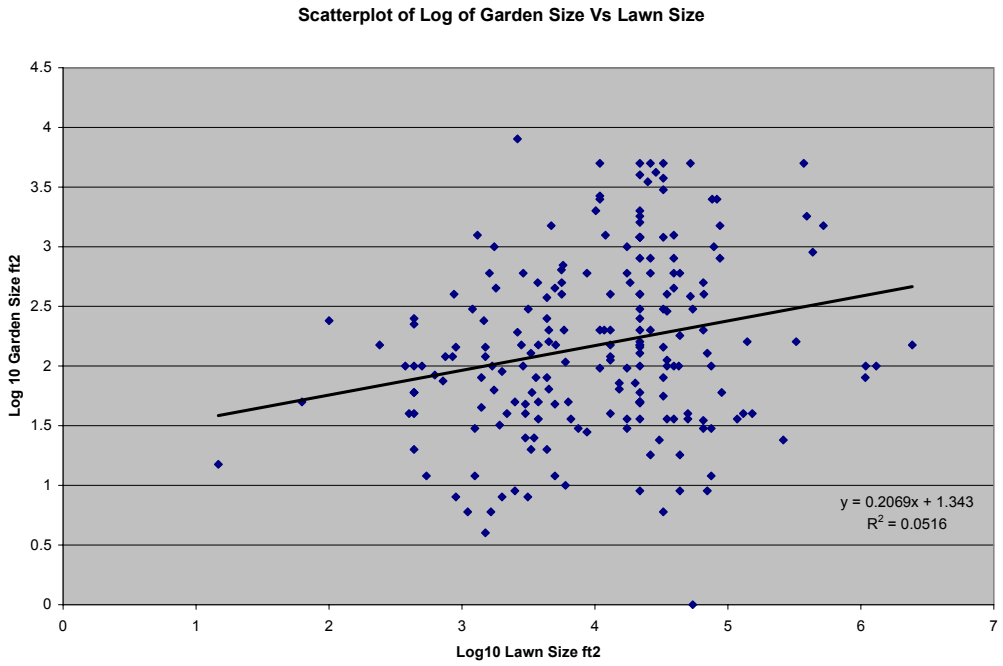
FS is the size of the finished space.

4. The portion of the yard that is covered by turf is assumed to be less than ten acres. Thus if the estimate of the yard is grater than 10 acres the model will assume that 10 acres are turf and the remaining portion of the lot is not managed as turf but is covered by, fields, woods, pasture, etc.

The presence of certain plants in the home are assigned based on data from NHGPUS (See Chapter 6). When a record is pulled for a home that record determines whether the home one or more of the following fruit trees, nut trees, or grape arbor, or whether the home has a vegetable garden.

1. Ornamentals are assumed to occur all homes. The model does not define the area of the ornamental plantings in the home.
2. The National Garden Survey (NGS) data indicates that gardens sizes in the surveyed homes had gardens that ranged in size from 0.1 m^2 to 740 m^2. Data from the NGS indicates that the size of the garden is independent of the size of lawn. Figure 1 shows a log-log plot of garden size vs. lawn size (R^2 of only 0.05). In addition, the survey reported that gardens occur even when there is no lawn). Finally, the garden size is independent of the presence of a lawn (mean size of all gardens is 569 ft^2 and the mean size of gardens in homes without lawns is 496 ft^2.

Figure 3-4



Based on this data, the model will assume that garden size is independent of lawn size. If a garden is determined to exist, then the model will randomly select a garden size from the sizes reported in the NGS. The distribution of sizes is given by the following cumulative distribution.

Table 3.5 Cumulative Distribution of Garden Sizes

Percentile	Garden Size (m^2)
0	0.0929
0.25	3.716
0.5	11.148
0.75	37.16
0.9	139.35
0.95	325.15
0.975	464.5
0.995	464.5
1	743.2

If a garden is created, the lawn size should be reduced based on the following equation:

Corrected Lawn size = Original Lawn Size- Garden Size

3.3.4 AIR-EXCHANGE RATES

Murry and Burmaster (1996) have published distributions of residential air exchange rates as a function of Census region and season. For each residence, a percentile value was randomly selected. Appropriate seasonal values were then selected from these published distributions. The seasonal values selected for each home were linked. This linkage is based on the assumption that houses with high air-exchange rates in one season will have high exchange rates in other seasons. Therefore, the model assigns a single percentile to a home and uses that percentile to select from each seasonal distribution.

For example, if the fifth percentile is assigned to the residence, the fifth percentile value for each season-specific distribution in that region is assigned to the residence for each of the corresponding seasons.

With this approach, there is no influence of setting, number of units,[39] or residence size (area or number of rooms) on the air exchange rate. Data to assess the influence of these factors were not located.

3.4 Types of Residences Considered in the Model

The current model is restricted to primary residences and to homes that are single detached or attached (one or more units). Thus, no second (vacation) homes and no vacant homes are included. In addition, mobile homes, tents, or "other" types of homes are not included in the model. Finally, the model will be limited to homes that have at least one kitchen, one bath, and one bedroom. In addition, as noted above, neither institutional residences (dorms, barracks, jails, nursing homes, schools, etc.) nor seasonal farm workers (and other temporary homes) are included in this version of LifeLine™. If an individual spent all or

[39] *A priori*, there is no reason to expect an influence of tenure on air exchange.

part of their life in a residence such as those excluded in this model, their exposure from pesticides used in those dwellings would not be included in the exposure calculations in this model. LifeLine™ can be modified to include these living scenarios. That would require adding data about the dwellings, people's mobility and residence duration, and activity patterns within the dwellings.

The model does include rural residences that will include farms. However, the model does not consider AI "track in" from agricultural uses of AIs or from individuals involved in commercial agricultural activities on their properties.

Finally, the model does not consider out-of-door exposures to AIs on the common portions of a condominium or apartment property (such as the pool, playground, or common spaces). The reason for this is that the NHGPUS survey results do not include data on AIs applied on such areas.

3.5 Golfing

Golfing results in recreational exposures to pesticides used on golf courses for older children and adults. These golfing exposures do not occur at home[40] but are influenced by the region in which the individual resides. Therefore, golfing is treated as a "residential" characteristic. In golfing an individual is assumed to golf on one or more courses in his or her region.

The golf course is divided into two areas: 1) greens and tees and 2) fairways and rough. Pesticides may be applied to either greens and tees or to greens, tees, fairways, and rough. Each golf course is assumed to have a single residue on the turf of either greens and tees or to greens, tees, fairways, and rough.

[40] Individuals may have putting greens or practice golf at home; however this is not addressed in this assessment since the focues is on exposures that happen in addition to residential exposures.

3.6 References

Bureau of the Census, U.S. Department of Commerce, and Bureau of Labor Statistics, U.S. Department of Labor. (1999) Current Population Survey, Annual Demographic Survey - March Supplement, 1992-1994 and 1996-1999 Data Sets.

Bureau of the Census and U.S. Department of Housing and Urban Development. (1993) American Housing Survey.

Murray, D.M. and D.E. Burmaster, "Residential Air Exchange Rates in the United States: Empirical and Estimated Parametric Distributions by Season and Climatic Region," *Risk Anal.*, 15(4): 459-465(1995).

National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services. (1998) 1996 Natality Data Set.

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) *National Home and Garden Pesticide use Survey*. Prepared by the Research Triangle Institute for the Office of Pesticides and Toxic Substances, Biological and Economic Analysis Branch.

CHAPTER 4. EXPOSURE-RELATED BEHAVIORS

Exposures to AIs are determined based on 1) information on the level of an AI in the environment and 2) a behavior that brings the individual into contact with that residue. This chapter addresses the ways in which LifeLine™ models behaviors that may bring a person into contact with residues in food, in the residential environment, or in tapwater.

The LifeLine™ system addresses three fundamental classes of activity; diet (the type and amount of food consumed), activities (the duration, location, and general description of behaviors) defined by the National Human Activity Pattern Survey (NHAPS) daily activity records, and activities not tracked in NHAPS. The dietary activities are determined by the selection of an appropriate record from the Continuing Survey of Food Intake by Individuals (CSFII) conducted by US DA. In this version, the 1989-1991 survey data are used. The second by the selection of an appropriate record from the NHAPS, conducted by EPA. The third group includes activities that are either not tracked by NHAPS or are not appropriately modeled using records. In Version 2.0, the activities that fall into this category are:

1. Interacting with pets;
2. Playing golf; and
3. Gardening.

4.1 Daily Activities

There are two steps for determining the appropriate inputs for the contact with residues in residential microenvironments and/or tapwater. First, an appropriate record is selected for the individual from NHAPS. Second, quantitative contact parameters are determined for each activity in that record, reflecting the user's specification of activity-specific parameters appropriate to the age of the modeled individual. This provides the level of quantitative detail needed to assess the exposures (mass of contaminate that reaches an individual) that would be expected to result from performing the activity for a duration time, given the residues in the microenvironment where the activity occurred.

Once these record and contact values are assigned, LifeLine™ uses simple equations to estimate the exposures or doses for each activity in a given day by way of the dermal, inhalation, and oral routes. These estimates of dose or exposure are then summed to give the total exposure or dose received on a given day.

4.1.1 NATIONAL HUMAN ACTIVITY PATTERN SURVEY

On any given day, a single NHAPS record is used to represent each modeled individual's activity patterns. It is essential to select an appropriate record for a given individual at a particular stage in life.

4.1.1.1 Record Selection

Selecting an appropriate record for a given individual on a given day requires distinguishing those records collected from individuals under circumstances that are similar to the person modeled. In other words, "binning" of the records to be sampled must precede sampling. Records from the appropriate bin for an individual at the time of sampling are eligible for sampling, while those in other bins are not.

There is, for any achievable survey, a trade-off between the precision of binning and the need to maintain an adequate bin size. Ideally, the definition of similarity (i.e., of the bin) should be made as strict as possible with regard to any variable that is likely to influence activity patterns. Setting up distinct bins for every age, sex, region, season, house type and setting in NHAPS, however, would result in bins with only a few records. Accordingly, LifeLine™ uses bins that reflect only those criteria that have been shown to greatly affect the patterns of daily activity. These factors are:

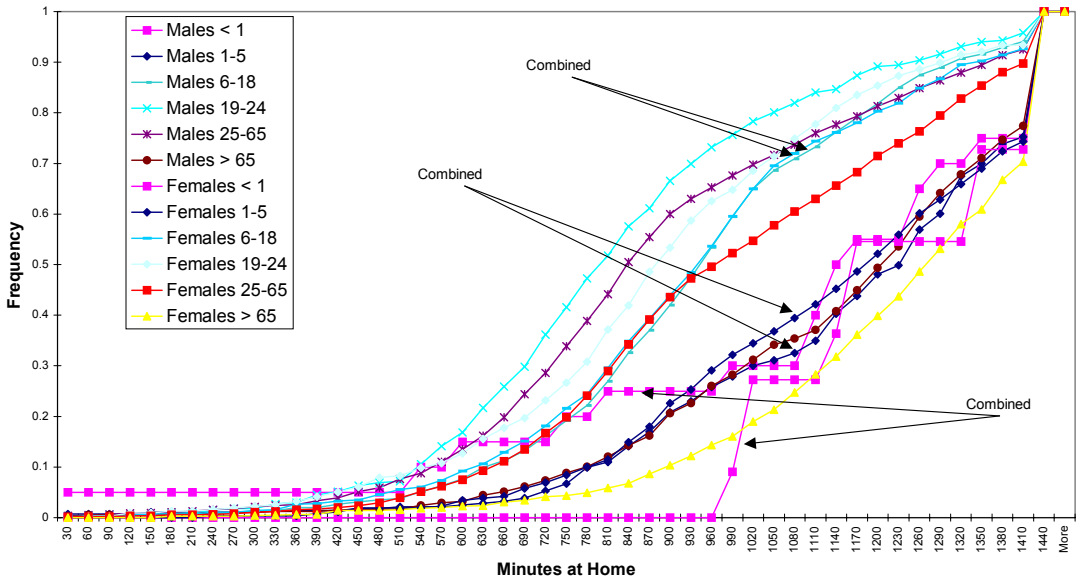
- Age;
- Sex;
- Season;
- Region;
- Day of the week (weekday or weekend); and
- Residence type (single family versus multifamily).

The following sections discuss each of these options. In each case, the NHAPS data were analyzed to identify actual differences in patterns between bins, to supplement *a priori* reasoning about the expected patterns in the data.

4.1.1.1.1 Age and Sex (9 bins)

There are a number of stages in individuals' lives where the residential activities are believed to be relatively stable. These include acquiring motor skills, reduction in the need for oversight, beginning education, and working outside of a home. These factors suggest that reasonable breakpoints for activities are <1, 2-5, 6-18, 19-24, 24-65, and >65. Figure 4-1 presents the distributions of time spent at home for these different categories by sex. Sex differences are greatest after schooling and child raising years, therefore, sex differences were considered for ages 19-24, 24-65, and >65.

Figure 4-1: Time Spent at Home (Male/Female)



4.1.1.1.2 Season and Region (3 bins)

Seasons affect activity patterns in two ways. Increasing temperatures result in more time spent out of doors. In addition, summer vacations affect the activities of both children and parents of children. On a year-round basis, regional differences in activity patterns are believed to be relatively small (EPA, 1997 EFH). However, regions do affect the number of days with warm temperature. Accordingly, LifeLine™ divides the regions and seasons into three categories referred to as SR categories. The three SR categories are, summer (all

regions), seasons in regions other than summer with warmer temperatures, and seasons in regions with cooler temperatures. This approach is somewhat cumbersome since certain regions, such as the West, include widely different climates.

Figures 4-2 and 4-3 present the cumulative distribution of time spent at home by region and season for weekends and weekdays. As the figures demonstrate, the three SR categories tend to track very closely for total time spent at home on both weekday and weekends.

Figure 4-2. Time Spent at Home on a Weekday by Season and Region

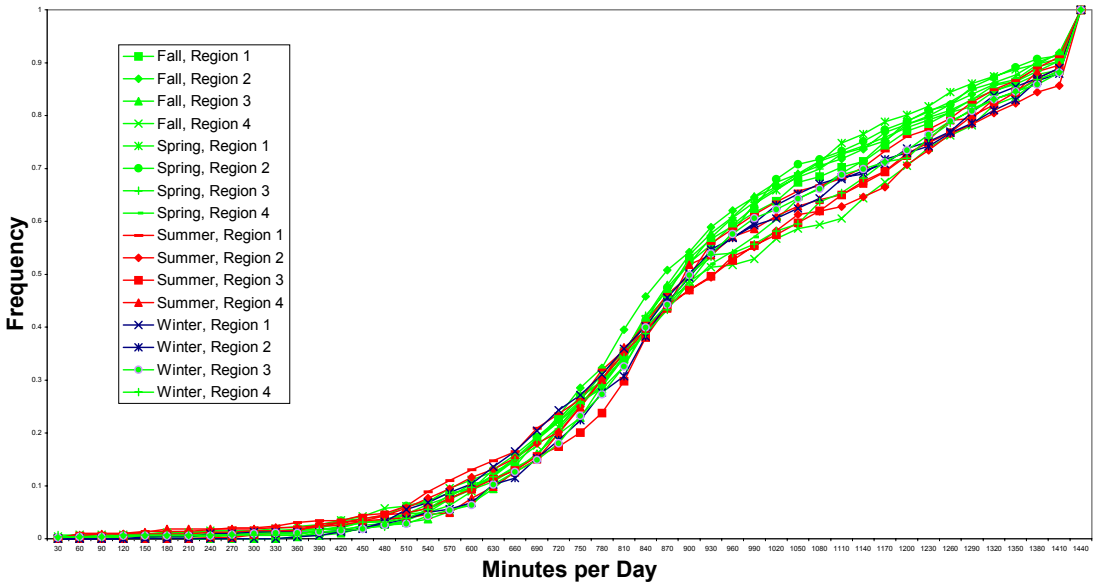
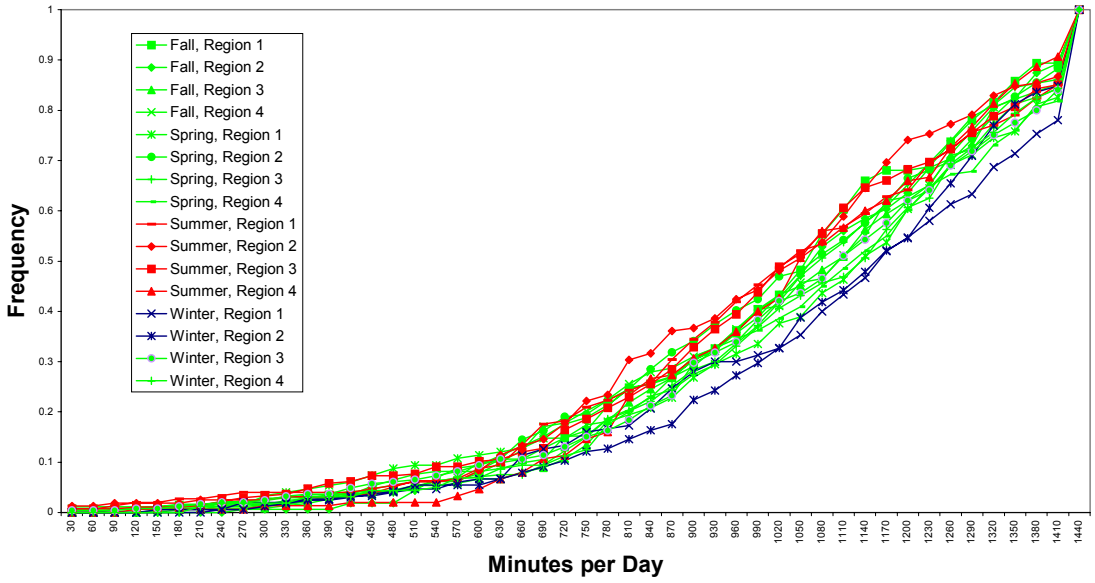


Figure 4-3. Time Spent at Home on a Weekend by Season and Region



4.1.1.1.3 Day of the Week (2 bins)

Weekends clearly differ from weekdays for individuals who work or attend school. Therefore, two records of weekends and weekdays are binned separately. As indicated in Figures 4-2 and 4-3, the activity patterns for these two types of days are different (based on total time spent at home).

4.1.1.1.4 Single and Multifamily Homes (2 bins)

Housing determines the number and nature of rooms. Therefore, single and multiple family residences will be binned separately.

4.1.1.2 Resulting Bin Size

Based upon the combination of *a priori* and empirical analysis of patterns in NHAPS, the individual records have been classified into 108 bins. Whenever an activity record is selected for an individual, it is drawn from the corresponding bin.

4.1.2 CONTENTS OF AN NHAPS RECORD

Each NHAPS record contains a time-series of activities and locations that begins at midnight and continues for 24 hours. The record includes starting and ending times of activities, and the location of those activities. For example, the following record lists

locations and activities for an individual on a fall weekend. No check is made on the apparent reasonableness of reported combinations of activities and locations.

Table 4-1. Example of Data from an NHAPS Record

Time (0-24:00 hours)	Location	Activity
0:00-10:00	BEDROOM	Sleeping/napping
10:00-10:30	KITCHEN	Eating
10:30-20:30	*Not in home*	
20:30-21:15	LIVING ROOM	Watching TV
21:15-22:00	DEN	Watching TV
22:00-24:00	BATHROOM	Sleeping/napping

4.1.2.1 Daily Aggregation of the NHAPS Record

The model makes the simplifying assumption that the time spent in a microenvironment is a single block of time. In reality and the original NHAPS record, the time may represent multiple episodes of time spent in a location. In this version of the model, all occurrences of a given activity in a particular location are summed to yield a total duration of that activity in that location. Future versions may model the actual time periods of individuals on a given day.

4.1.2.2 Time Spent Away from the Residence

Activities outside the residence are not retained from the NHAPS record in this version.

4.1.3 SPECIAL CONSIDERATIONS IN MATCHING NHAPS RECORDS TO CURRENT RESIDENCE

The construction of a model of residences requires data that have been collected by a number of surveys NHAPS, NHGPUS (See Chapter 5), and AHS (See Chapter 3). In a number of cases, the microenvironments evaluated in the different surveys do not match. In order to use the data from these surveys in a single model we constructed the following rules.

4.1.3.1 Living Room/Family Room/Den

The NHAPS identified the residential microenvironment of Living Room/Family Room/Den. In many homes, the AHS identifies both a den and a living room. Where both

rooms are identified, the model randomly assigns the activity to either room (for purposes of assigning room size and tracking residues) on any given day. Where a home only has one of these rooms, the model assumes that all of the activities happen in that location.

4.1.3.2 Locations in Small Residences

As described above, NHAPS records have been binned using a limited number of characteristics. As a result, a NHAPS record selected for an individual can indicate that he or she spent time in a residential location that is inconsistent with the characteristics of her residence. For example a record may specify that time is spent in a home office when the home may not have such a room.

This problem is addressed in three ways. First, all records of homes taken from the AHS have at least one kitchen, one bedroom, and one bathroom. Thus, the locations where individuals spend the majority of time and the locations where pesticides are most often used are always addressed in the model. Second, as noted above, binning links the activity records with the type of home. If a person resides in an apartment, NHAPS records that include garages or basements are unlikely to be selected. Third, when locations defined by the NHAPS (living room/family room/den, dining room, bathroom, bedroom, study/office, garage, basement, utility room/laundry room) do not occur in the home, the time/behaviors spent in these locations are allocated to other rooms. Table 4-2 presents the rules that will be used to reassign locations for an individual's daily activity pattern record into existing rooms of a residence.

Table 4-2. Rules for Substitution of Microenvironments			
Microenvironment	First Substitution	Second Substitution	Third Substitution
Cellar	Other	Kitchen	
Bedroom	*Always there*		
Bathroom	*Always there*		
Half Bath	Bathroom		
Kitchen	*Always there*		
Living room	Den	Other	Kitchen
Dining room	Kitchen		
Den	Living room	Other	Kitchen
Other	Kitchen		
Halls/Stairs/Closets	Other	Kitchen	
Garage	Other outdoors		
Other Outdoors	*Always there*		
Office	Other	Living room	Kitchen

4.1.4 QUANTITATIVE PARAMETERS

As can be seen above, the activities reported in NHAPS do not provide sufficient information to support a quantitative assessment of exposure. In order to carry out such an assessment, the user must supply detailed estimates of the values of key exposure-related parameters for each behavior[41]. A set of "preset" values for these parameters has been provided with LifeLine™. In addition, the *Activity Description* program allows the user to modify the presets or provide completely new descriptions of each activity. This includes both the activities from NHAPS records and the non-NHAPS activities, playing with pets, gardening, and golfing.

The following exposure parameters require quantitative values (in some cases zero) for each of the NHAPS activities:

[41] In actuality, the values for these parameters will vary from event to event and for different individuals. Thus, the values cannot be fully described by a single value. Future versions of LifeLine™ will allow the user to input distributions for the parameter values that reflect inter and intra-individual variation in the values of these parameters.

- A general measure of activity level that is used in the determination of inhalation rates;
- A dermal transfer coefficient that has been normalized to surface area of the individual;
- A clothing protective factor;
- The fraction of the hand that is placed in the mouth;
- The refreshment rate for residues on the hand;
- Soil intake; and
- Grass intake.

For each activity, the user may specify different values of the parameter for different ages. The specific ages at which parameter values change can be independently specified for each activity.

Version 2.0 of LifeLine™ comes with a "preset" of values for these exposure parameters contained in the file *preliminary assumptions.acd*. The LifeLine™ development team developed this set of values. The values are intended to be reasonable but conservative (i.e., more likely to over estimate exposures than under estimate exposures) of the values of the parameters for typical individuals. The values have been loosely based on parameter values used in the Draft SOPs (EPA, 1999) and other agency documents (EPA, 1997).

While these values reflect discussions with EPA staff, they are the product of the LifeLine™ development team and do not express any formal position by the Environmental Protection Agency. The values are intended to be a starting place for users to develop their own sets of assumptions and to facilitate users learning to use the software. The development team takes no position on the appropriateness of these values for any specific assessment. The user should carefully review these values and replace them with alternative values that reflect the availability of new data or the specific goals of their assessments.

4.1.4.1 Activity Level

This is an assignment of activity at different ages to one of five categories of energy expenditure, and is used in the calculation of the individual's corresponding respiration rate. The five categories include:

- Rest,
- Sedentary,
- Light,
- Moderate, and
- Heavy.

The values in the initial preset are derived from Funk et al, 1998. This determination is used in the characterization of inhalation rates. Each class has a corresponding multiplier that is used to adjust the individual's specific baseline inhalation rate (See Chapter 2).

4.1.4.2 Dermal Transfer Coefficient (Normalized) (hr^{-1})

Dermal exposure to AIs is usually assessed using dermal transfer coefficients. A dermal transfer coefficient is defined as the mass of an AI that reaches the skin of an individual performing an activity for a specified time divided by the dislodgeable mass of AI in a specified area of a surface (typically a floor). The units of the dermal transfer coefficient are cm^2/hr. The values of the coefficient have been determined from studies of adult dermal exposures (EPA, 1999).

The values of the dermal transfer coefficients are believed to be proportional to the surface area of the individual. That is, a person whose hand has a surface area twice that of another individual will have twice the dermal exposure.[42]

As discussed in Chapter 2, LifeLine™ calculates the surface area of each modeled individual. The model takes advantage of this data to more accurately estimate the dermal exposure of each individual. The approach used to achieve this is to create a new variable the normalized dermal transfer coefficient, T_C'.

[42] This assumption is the basis for the extrapolation of transfer factors from adults to children.

T_C' is defined as the dermal transfer coefficient (T_C) observed in a study of dermal exposures with adults divided by the surface area of a typical adult (20,000 cm^2).

$$Tc' = Tc / SA$$

The values of Tc' in the preset were determined in a two-step process. First, each activity was categorized into four categories, no exposure, low, moderate, and high dermal exposure to floors and objects with large surfaces (furniture, tables, etc.). Second, the four categories were linked to the following set of values for dermal transfer coefficients for adults.

Table 4-3. Dermal Transfer Coefficients for Various Levels of Activity	
Category	**Dermal Transfer Coefficient (Tc) (cm^2/hr)**
None	0
Low	100
Moderate	1,000
High	43,000

These values were normalized by dividing by the surface area of an adult (20,000 cm^2).

Table 4-4. Normalized Dermal Transfer Coefficients for Various Levels of Activity	
Category	**Dermal Transfer Coefficient (Tc') (cm^2/hr)**
None	0
Low	0.005
Moderate	0.05
High	2.2

The model takes the value of T_C' and calculates a person-specific value of T_C by multiplying by the individual's surface area.

4.1.4.3 Frequency of Hand-To-Mouth Events (h^{-1})

This is the frequency with which some portion of the individual's hand is placed in the mouth during an activity. The preset values for this factor follow the agency's default assumptions of 20 events per hour for children aged 0 to 6 six years. Smaller numbers of events are used for older ages. The user can change this assumption.

4.1.4.4 Fraction of Hand Placed in the Mouth (unitless)

The default value is 0.114 for most activities. This value was estimated based on the assumption that the typical event will consist of the individual placing two fingers in the mouth. A smaller fraction of the hand is used for older ages. The user can change this assumption.

4.1.4.5 Replenishment Rate (unitless)

The replenishment rate is a measure of the ratio of the average dislodgeable residue in the hand to the amount that would occur immediately after contact with a surface. This value would be expected to be less than 1.0 due to factors such as:

- Losses from contact with other surfaces prior to a mouthing event,
- Touching a surface with only part of the hand, or
- The failure to re-contact a surface in between mouthing events.

The replenishing rate is likely to be affected by the activity. Where a child is sleeping or napping, the hand may be frequently placed in the mouth, but there is no opportunity for replenishment from a carpet or other surface with pesticide residues. Similar reductions may occur for eating, bathing or other activities that would keep the child from interacting with exposure sources.

The replenishment rate does not reflect the binding of the residue to the skin. This factor is accounted for in the extraction efficiency term (See Chapter 5).

The preset value for this term is 1.0 except for activities such as sleep/napping and sedentary activities, where a value of 0.1 is used. This lower value reflects the limited potential for re-contact with surfaces. Lower replenishment rates are assigned for older ages reflecting the lower potential for interaction with surfaces in adults. The user can change this assumption.

4.1.4.6 Grass/Vegetation Consumption Rate (cm^2/h)

The amount of grass and vegetation that is consumed per hour is linked to the duration of time spent out of doors. The value used in the preset is based on the conservative

assumption that a small child consumes 25 cm^2 of grass during a two-hour period of contact with turf. The value of grass consumption is set as 25 cm^2/2h or 12.5 cm^2/h. This assumption is based on the values used in the EPA draft SOP. For older children and adults, smaller amounts of grass consumption are assumed (1 cm^2/hr). These rates are applied to those activities with high levels of dermal contact. The user can change this assumption.

4.1.4.7 Soil Consumption Rate (mg/hr)

The value for the soil consumption rate is 100 g/day and is taken from the 1999 draft SOPs[43]. The amount is assumed to be a function of time spent out of doors. As with the grass consumption rate, the default soil consumption rate is normalized to a 2-hour period or 50 mg/hr. Following EPA guidance, the soil consumption rate for children and adults over the age of 6 is assumed to be 1/2 that of children. These rates are applied to those outdoor activities with high levels of dermal contact. The user can change this assumption.

4.2 Other Activities

This group of activities includes activities that are not tracked by the NHAPS. The NHAPS record establishes a 24-hour period and assigns an activity to each minute of the day. As discussed above, this assignment of activities is used to estimate post-application exposures. The activities in this section do not fit into this framework of time. In Version 2.0 these activities happen in addition to or in the case of pets simultaneously with the NHAPS defined activities. This causes a problem in that certain individuals are modeled as having more than 24 hours of activities in a day. This issue is not likely to cause a significant overestimation in dose but should be addressed in future version of the model. One approach to do this would be to replace the NHAPS defined times at a specific block of time with the time spent in the activities.

[43] Soil consumption is a complex behavior and can occur from ingestion of house dust that is contaminated by soil tracked into homes. Version 2.0 does not consider track in of pesticides. Future versions of LifeLine™ may include this pathway.

4.2.1 GARDENING

LifeLine™ Version 2.0 allows the user to enter data on the frequency and duration of gardening events. These data are entered in the AIPD. These data are used if the person is residing in a home with a garden. This period-of-time is assumed to happen in addition to the NHAPS defined activities.

4.2.2 GOLF

4.2.2.1 Background

The probability of golfing is determined in LifeLine™ based on published data on the demographics of golf (NGF, 2001). These data suggest the following conclusions. First, golfing, defined as playing a complete round of golf, does not appear to occur with any regular frequency in children under the age of 12. Second, the probability of playing golf is not equivalent across different socioeconomic groups, ages, and genders. Golfers tend to be college-educated, older, and male (NGF, 2001). Third, golfers can be defined in terms of four categories:

- Junior,
- Occasional,
- Moderate, and
- Avid golfers.

Using this data Version 2.0 defines each individual based on income quartile, gender, and age.

The probability of playing golf is modeled in a series of steps:

1. Is the person's age 12? If not skip this module until they are 12.
2. Is the person a junior golfer? If not skip until 18.
3. Is the person 18?
4. Is the person a junior golfer? If yes then assign them to a golf category (occasional, moderate, or avid). If not, then did the person become a golfer? If yes, assign them to a golfing category (occasional, moderate, or avid). If not a

junior golfer and did not become a golfer, then skip this module until the next person.

5. Given the type of golfer and their gender, determine the frequency of play (rounds per year).
6. Given the frequency of play, what is the daily probability of play.

All golfers between the ages of 12 and 18 are defined as junior golfers. The probability of becoming a junior golfer at age 12 is a function of the gender and the level of education of the head of the household. The current model of LifeLine™ defines each individual in terms of the education of the individual's mother. (These data are assumed to be predictive of the education of the head of the household and the modeled individual when the individual reaches age 18). The probabilities of being a junior golfer are given in the following table:

Table 4-5. Probability of Becoming a Junior Golfer At Age 12

Gender	SES Quartile	Probability
Male	1	0.037
Male	2	0.054
Male	3	0.065
Male	4	0.099
Female	1	0.008
Female	2	0.012
Female	3	0.015
Female	4	0.023

Data from the NGF survey suggests that the fraction of junior golfers is relatively constant from ages 12-17. As a result, the model will assume that once becoming a junior golfer the individual will remain a junior golfer.

At age 18, all junior golfers are moved to one of three adult categories of golfing occasional, moderate, and avid. The probability of becoming one of the three categories

will be different for each of the eight combinations of gender and income. The following table gives the probabilities of becoming each of the adult categories. Notice males tend to go to the avid category in larger numbers than females junior golfers.

Table 4-6. Probability of a Junior Golfer Becoming One of the Three Adult Categories of Golfers

Gender	SES Quartile	Occasional	Moderate	Avid
Male	1	0.422	0.296	0.282
Male	2	0.389	0.304	0.307
Male	3	0.389	0.335	0.276
Male	4	0.371	0.339	0.290
Female	1	0.629	0.194	0.176
Female	2	0.609	0.200	0.192
Female	3	0.608	0.219	0.173
Female	4	0.597	0.222	0.181

The number of adult golfers is larger than the number of junior golfer to provide these additional golfers, a small number of non-golfers have to become adult golfers. Thus at age 18 the non-golfers are assigned to one of the three golfing categories based on the gender and their SES category.

Table 4-7. Probability of Non-Golfer Becoming a Golfer at Age 18				
Gender	**SES Quartile**	**Occasional**	**Moderate**	**Avid**
Male	1	0.0060	0.0042	0.0040
Male	2	0.0016	0.0013	0.0013
Male	3	0.0068	0.0058	0.0048
Male	4	0.0118	0.0108	0.0092
Female	1	0.0042	0.0013	0.0012
Female	2	0.0030	0.0010	0.0009
Female	3	0.0058	0.0021	0.0017
Female	4	0.0097	0.0036	0.0029

Using the above approach the model will be able to determine if an individual is a non-golfer or a junior, occasional, moderate, or avid golfer on each year of their lives. This assignment will reflect the income and gender of the individual but may not capture racial differences.

The golfers in these categories have different frequencies of playing golf. Based on the NGF survey:

Table 4-8. Rounds per Year		
	Male	Female
Junior	15.4	9.9
Occasional	3.2	2.8
Moderate	14	13.7
Avid	68	63.7

Once the assignment of the golfing categories is established the frequency reported by category and gender can be taken from Table 4-8. The daily probability of playing golf is estimated based on the following formula:

If the day falls in the warm portion of the year:

$$DP_G = A_G / (FY * 365)$$

If the day falls in the cool portion of the year:

$$DP_G = 0$$

Where,

DP_G is the probability of going golfing on a given day.

A_G is reported number of games played for the persons golf category.

FY is the fraction of the year that is warm.

Games can be played on consecutive days.

4.2.3 PESTICIDE APPLICATION

The method of estimating applicator doses (unit transfers) does not require the estimate of duration. However, conceptually an application of pesticides would take time and this time would be in addition to the NHAPS defined activities. Future versions of LifeLine will assign time of the day and durations for this activity.

4.2.4 PLAYING WITH PETS

Individuals interact with pets in multiple locations in a series of encounters that last for varying lengths of time. LifeLine Version 2.0 allows the user to inter a single estimate of the total duration of time playing with pets. This value is entered on the user's preference page. The interactions with pets are assumed to occur at the same time as the NHAPS defined activities. For example, a cat or dog may be petted while the person is performing activities such as watching television, or reading. Therefore, the time playing with pets is not a duplication of the NHAPS activities.

Individuals interact

4.3 Daily Diet

In general terms, dietary activities are sampled according to the same procedures as other activities. The binning criteria for dietary records are, of course, different, but the logical process for identifying bin boundaries is the same. Also, because dietary records provide an adequate level of detail for exposure assessment, given residues in foods (See Section 5.2), no further specification, such as that using user-specified parameters for other behaviors, is required.

4.3.1 DIETARY INTAKE DATA

LifeLine™ Version 2.0 provides the user with a choice of two versions of the USDA's Continuing Survey of Food Intakes by Individuals (CSFII). As with Version 1.0, the user may select the 1989-1991 CSFII, with associated food recipe files. Version 2.0 also, however, allows the user to select the 1994-1996, 1998 CSFII, with associated translation files developed by EPA and USDA. Section 5.1 contains a more detailed discussion of the differences between the new translation files and earlier recipe files.

4.3.1.1 General Characteristics of CSFII

While the newer CSFII data represent a continuation of the earlier surveys, the differences between the surveys are nearly as marked as are the commonalties. The user should carefully consult the documentation provided with CSFII data in order to fully consider these differences.

This series of surveys conducted by USDA used a stratified area probability sample of individuals residing in households in the U.S. Households represented a cross section of the population of the 48 conterminous states and the District of Columbia, although low-income households were intentionally over-sampled. Interviewing took place over the entire year for each year of the survey. The surveys obtained reports of dietary intake (food and beverages consumed both at home and away from home) of all individuals in survey households for the survey days.

Households and individuals were surveyed in all four seasons and on all days of the week. In addition to information on food consumption, the survey collected physiological and demographic data such as sex, age, self-reported height and weight, ethnic group, pregnancy and lactation status, and household income. This information permits sorting and aggregation of data, as well as tags for bridging to other data sets within LifeLine™. The survey samples included nursing infants, but consumption of breast milk was not estimated in the survey.

Food intake was recorded by time of day and by eating occasion (breakfast, brunch, lunch, dinner, supper and snacks) as defined by the respondent. Separate entries were made in the database for each food consumed. Quantities of foods and beverages consumed were recorded in household measures, weights, dimensions or common units (e.g., slice), and then were converted to grams by USDA.

4.3.1.2 Features of the 1989-1991 Survey

This survey addressed consumption over a three-day period for each sampled individual. The Day 1 individual intake was collected by trained interviewers using a 24-hour recall of foods consumed the previous day, while Day 2 and Day 3 intakes were based upon food intake records maintained by the respondent for the day of the interview and the following day.

Approximately 5,000 individuals residing in over 2,000 households participated in each year of the 3-year survey. Although the majority of individuals reported consumption for all three days of the survey, some individuals reported consumption for only one or two days. Weights were developed by USDA to adjust for over- and under- representation of certain population subgroups in the un-weighted sample due to the sample design, non-response, and unequal interviewing across seasons and days of the week.

4.3.1.3 Changes for the 1994-1996, 1998 CSFII

Two fundamental changes in survey design are most noteworthy. First, rather than collecting data on three sequential days from each individual, the survey explicitly collected data on two *non-sequential* days for each individual. Thus, whatever temporal

dependencies exist in the earlier survey (as opposed to patterns of consumption that are characteristic of a sampled individual over a prolonged period), different day-to-day patterns should be seen in the more recent survey.

Second, because the basic survey design was found to collect less data than desired on the diets of children, in 1998 a focused survey of *children's* diets was conducted. This was designed in such a way as to permit a combination of these data with those collected in the general survey between 1994 and 1996, and results in a far richer sample of children's diets than was available in earlier surveys. For example, the 1989-1991 survey included records from 204 different individuals between the ages of 1 and 4. The 1994-1996, 1998 survey, in contrast, contained records for 5,886 children in this age range.

4.3.2 DETERMINATION OF DIETARY PATTERNS AND DATA BINS

4.3.2.1 Previous Binning Procedures

Traditionally, the CSFII data have been used as a pool from which individual records were selected using Monte Carlo techniques. The pool of records could be used as a single pool, but was more likely segregated into "bins" using some separation criteria. In previous dietary assessment software tools used by EPA, these separation criteria were devised in a very informal (almost intuitive) way. Professional discretion was used to describe 21 population subgroups. These were based on informal patterning exercises. The logic for these groupings were:

- Regional: there were four defined regions in the design of the USDA CSFII survey. These regional categories were preserved primarily because one region—Western U.S.- was populated primarily by the State of California, which had many of its own risk assessment regulatory programs and thus would find this regional category useful. More refined geographical subdivision was not possible.
- Seasonal: The four seasons were maintained because they also were integral to the food consumption survey, and could be useful in revealing exposure situations which were unique because of seasonal changes in consumption, especially for some fruits and vegetables.
- Ethnic: It was assumed that ethnicity could define unique eating patterns. No formal investigation was made to substantiate this, nor was it equated with socioeconomic status or regional relationships. Ethnicity was defined as it appeared in the survey

questions. No attempt was made to define these patterns before including them as population subgroups of interest.

- Age: It was assumed that infants would eat differently than young children, who would eat differently than toddlers, who would eat differently than older children, etc. The age groupings were created "intuitively" from professional judgment, but no formal patterning was explored. To some degree, the age grouping reflected practical grouping of records so that adequate numbers of records were contained in any one age grouping. The relatively small number of young children and infants in the survey made detailed subdivisions very difficult because small numbers of records would result for each bin. Age categories were also made to reflect the approximate onset of puberty in young men, and the approximate ages of childbearing for women.
- Nursing/non-nursing: It was assumed that these infants would have different eating habits.
- Pregnant/lactating: It was assumed that these women might have different eating habits.

Thus, the traditional population subgroups were professionally well-considered and created with some concepts of patterning, although the process was informal. The patterns were thought to reflect:

- Dietary profile differences
- Physiologically unique situations such as pregnancy, onset of puberty, or lactation. Notably, physiological changes associated with aging were not considered.
- Possible differences in food forms of the foods eaten in different age groups
- Practical considerations of numbers of respondent records available in the grouping
- Intrinsic survey design parameters (such as region and season)

Other subgroups could have been constructed from the survey records, such as

- Different age groupings,
- Vegetarians, or persons with medically-influenced diets. The definitions for "vegetarian" were difficult to resolve and the idea was thus abandoned. Similar difficulties for medically-influenced diets led to the rejection of this category, also.
- Socioeconomic subgroups, defined by income, education or other relevant parameters. Until very recently, this parameter was controversial or considered inappropriate for a separate assessment category.

Binning Procedures for Dietary Data in LifeLine™ Version 2.0

The CSFII data were examined to find dietary profiles that may be unique, and to justify groupings of respondent records based on explicitly identified similarities. Each day of food records was treated as independent, even though in the 1989-91 survey each participant provided up to three sequential days of dietary records, and in the 1994-1996, 1998 survey, each respondent may have provided two (non-sequential) days of reporting.

The effort to identify patterns of dietary behavior considered both the characteristics of the reporting individuals and the characteristics of reported behavior. Among the parameters of dietary profile that were examined for this purpose were:

- Body Mass Index (BMI),
- Caloric Intake (kiloCalories per day),
- Frequency of eating within a day—
  number of eating occasions per day,
  number of meals per day,
  number of non-meal eating occasions per day
- Characteristics of the eating occasions—number of foods per eating occasion
- Food selection overall—
  number of foods eaten per day,
  number of different foods eaten per day
- Mass of food eaten per day or per day per unit body weight

These parameters were examined together within a matrix of age, sex, and season. There were too few infants less than one year of age to allow analysis of trends in eating by month of age (particularly for the 1989-1991 survey), although it was apparent that significant differences do exist between early infancy and 9-12 month-olds.

An effort was made to find natural patterns in these dietary descriptors, such as a decrease in the number of meals or the increase in the number of different foods consumed in a day, from infancy to adulthood. Overall, there were surprisingly few differences observed between males and females of the same age (there are differences in total consumption that reflect differences in body weight). When taken together, these observations allowed the food records to be separated into age-based bins.

4.3.2.1.1 Bins for the 1989-1991 CSFII Data

For the 1989-1991 survey, the data supported the following bins based on age (seasons were binned separately within each age category):

Bin 1: Nursing infants less than one year

Bin 2: Non-nursing infants less than one year

Bin 3: Age 1-4 years

Bin 4: Age 5-12 years

Bin 5: Age 13-25 years

Bin 6: Age 26 and above

There were other observations that reinforced the decisions about the age limits of the bins (See Table 4-5). For example, there is an increase in the number of daily snacks for 1 to 4-year olds, an increase in the proportion of food records reporting lunch for 5 to 13-year olds, a decrease in the proportion of food records that included breakfast for young adults and an increase in the number of different foods eaten by those older than 25. Such patterns dictated the binning decisions.

Table 4-9. Summary of Key Characteristics Used in the Selection of "Bins" of Dietary Patterns for 1989-1991 CSFII

"Bin"	Mean E.O./day *	Range E.O./day	Mean # Foods/day	Range # Foods/day	Mean # Different Foods/day	Range # Different Foods/day
Nursing Age < 1	6.49	1-14	7.82	1-25	2.77	1-16
Non-Nursing <1	5.64	1-12	8.63	1-25	4.55	1-18
Age 1-4	4.30	1-17	11.99	1-29	10.23	1-26
Age 5-14	3.75	1-13	12.44	1-32	11.26	1-28
Age 15-25	3.50	1-16	11.10	1-52	9.97	1-28
Age 26+	3.75	1-25	13.22	1-50	11.57	1-37

***E.O. is Eating Occasion**

LifeLine™ permits sampling of seasonally-defined records from these bins.

4.3.2.1.2 Bins for the 1994-1996, 1998 CSFII Data

Because of the greater number of records for children, LifeLine™ was able to select finer bins for the 1994-1996, 1998 survey than for the 1989-1991 survey. The following age bins were used (again, separate sampling was employed for each season):

Bin 1: Nursing infants less than one year
Bin 2: Non-nursing infants less than one year
Bin 3: Age 1 year
Bin 4: Age 2 years
Bin 5: Age 3 years
Bin 6: Age 4 years
Bin 7: Age 5 years
Bin 8: Ages 6-7 years
Bin 9: Ages 8-11 years
Bin 10: Ages 12-14 years
Bin 11: Ages 15-25
Bin 12: Ages 26 and above

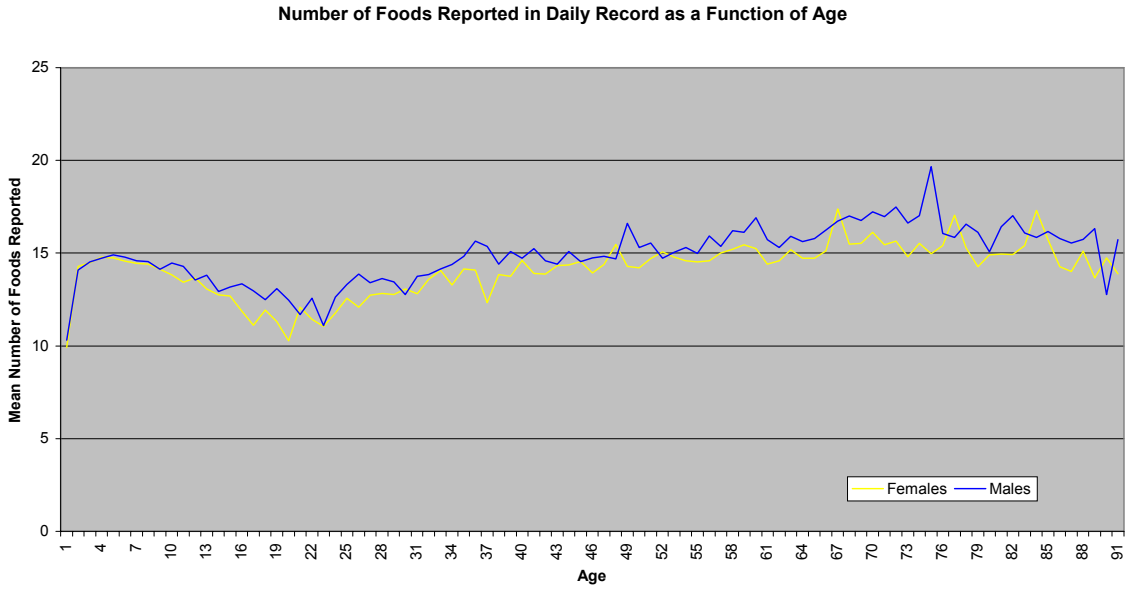
As in the case of the 1989-1991 data, there are clear trends in dietary patterns across these bins:

Table 4-10. Summary of Key Characteristics Used in the Selection of "Bins" of Dietary Patterns for 1994,98 CSFII

"Bin"	Mean E.O./day *	Range E.O./day	Mean # Foods/day	Range # Foods/day	Mean # Different Foods/day	Range # Different Foods/day
Age < 1	2.545	1-7	10.104	1-27	4.825	1-19
Age 1	4.031	1-7	14.187	1-34	11.400	1-29
Age 2	3.869	1-6	14.503	1-34	12.464	1-32
Age 3	3.878	1-7	14.749	1-37	12.833	1-33
Age 4	3.842	1-6	14.828	1-38	13.048	1-34
Age 5	3.779	1-6	14.683	2-34	13.049	2-27
Ages 6-7	3.714	1-6	14.511	4-35	12.993	3-31
Ages 8-11	3.599	1-6	13.946	1-38	12.639	1-32
Ages 12-14	3.424	1-5	13.060	2-52	11.833	2-42
Ages 15-25	3.234	1-6	12.186	1-44	11.042	1-32
Age 26+	3.456	1-7	14.991	1-56	13.281	1-47

With the smaller bins, the gradations between bins are smaller, reflecting continuous trends over age, as can be seen in the data for total number of foods, below:

Figure 4-4



4.3.2.2 Weighting

The CSFII data set contains data from independent samples drawn over three years (four in the case of the 1994-1996, 1998 data) from the basic sampling frame. For each survey, USDA provided sample weights with each record to be used with the full data set. These sampling weights account for survey non-response and to adjust for specified population totals. The sampling plan for the survey was based on household samples and an over-sampling of low-income populations. This satisfied the intent of the survey (i.e., assessing the nutritional status and possible nutritional deficits in these populations).[44]

4.3.2.2.1 Available Weights for CSFII Data

Separate weights have been provided for several general uses of the 1989-1991 data:

For household level analysis of the full data set across the three years;
For analysis of persons with day one intakes for all three years; and
For analysis of persons with three days of intakes across three years of records. (Note that a significant number of respondents did not provide all three days of the recording. These persons are excluded when this weight is applied.)

[44] A complete discussion of the methodology of the survey and the development of weights for the survey records are available from USDA, Food Surveys Research Group, Beltsville Human Nutrition Research Center, and Agricultural Research Service as part of the documentation for 1989-91 Continuing Survey of Food Intakes by Individuals and the Diet and Health Knowledge Survey, Section D of the Methodology Documentation.

A second set of weights, annual weights, is provided for these three analysis types if data for a given year are considered separately.

Similar options are provided for the 1994-1996, 1998 data, with the difference that there are four-year weights applicable to the entire data set, that reflect not only all years of data but also the fact that the 1998 data constitute a focused study on children.

It is important to note that the USDA weights are calculated to reflect the full survey dataset or the USDA-defined subset of data for specific weights. None of the available weights reflects the division of the data into sub-groups intended to represent specific distinct populations (binning) as defined in LifeLine™ or any other currently available exposure/risk analysis software.

The applicability of any of the available weights to a particular binning scheme depends upon the extent to which the weighting criteria that had been applied to the entire data set (or a particular subset of records) were uniformly applicable to the various bins that had been defined to segment those records. *A priori*, this is unlikely to be the case. The question becomes whether the advantages provided by using the weights offset the inevitable distortion involved in applying those weights to binned records.

The LifeLine™ user has the option (in *Analysis Preferences*) of opting to use weights or not when selecting CSFII records. The weights provided in Version 2.0 of LifeLine™ are from the three-year analysis of persons with three days of intakes for the 1989-1991 survey, and for the four-year analysis of persons with two days of records for the 1994-1996, 1998 data. There is no universally correct choice for this option, but these weights are arguably more appropriate for the data sets from which these bins were constructed.

As noted, the merits of applying weights greatly diminish for analyses that involve binning of subsets of the records. In general, it is suggested that analyses be conducted both with and without the application of sample weights to consider the possible differences in the exposure/risk profiles.

4.3.3 RECORD CONTENTS

The full CSFII records contain more information than needed by the present version of LifeLine™. Therefore, the system uses smaller file extracts of the needed information. The smaller of these extracts contains the information that allows the assignment of records to appropriate bins based on age range and season, and contains the weights for each record. The larger file contains the actual consumption data for the record, as well as key identifying information:

- Identifying data for CSFII (ID and Record Numbers);
- Day number and number of completed days for this person;
- Three-year, three-day record weight;
- For every food reported that day; and
  - ⇒ Food item code,
  - ⇒ Time of consumption, and
  - ⇒ Mass consumed.

4.3.4 RECORD USE

Version 2.0 of LifeLine™ uses a one-day time step for exposure analyses (future versions may employ finer gradations of temporal analysis). Accordingly, consumption values for each food item are summed for the entire day, and then matched with the appropriate food-specific residue value (See Chapter 5). For example, if a person ate 100 grams of apple pie[45] at 7:00 AM, 75 grams at noon, and 150 grams at 6:00 PM, the system would make a single calculation of the residue mass in 325 grams of apple pie.

For each food reported in the daily record, the LifeLine™ draws a single value for food-specific residue concentration on that day, and calculates the total residue mass ingested as a function of eating that food. Summing these values for all foods consumed on a day indicates the dietary exposure (i.e., residue mass ingested) for that day.

[45] For example, Food code 533010 is defined as PIE, APPLE (INCLUDE APPLE-PEACH, APPLE-BERRY)

4.4 Tapwater Consumption

CSFII records contain information on the consumption of water. Using these data, however, requires correction to control for the source of the water, as opposed to the location of consumption. Methods exist for estimating the intake of tap water and commercially added drinking water (as well as bottled water). The methods entail utilizing food technology data regarding the water content of foods as well as the dietary consumption surveys to detail the intake of water by individuals in the surveyed populations.

The EPA is actively pursuing the analysis of tapwater consumption, using the CSFII. Accordingly, the Development Team decided not to pursue the analyses required to accurately address tapwater consumption pending the outcome of those analyses. Instead, tapwater consumption values that reflect current EPA policies, as well as an interim analysis of data in the 1994-1996, 1998 CSFII were implemented in Version 2.0. These are likely to be replaced following the conclusion of EPA's analysis and publication of its associated policies.

4.4.1 WATER CONSUMPTION FOR THE 1989-1991 DATA

The EPA has established assumptions about water intake that are used across the Agency and form the basis of national drinking water standards (EPA, Exposure Factors Handbook, Vol. 1, General Factors, "August, 1997, Office of Research and Development, EPA/600/P-95/002Fa). These values reflect the estimation that a 10 kg child (age 12 months) consumes 1 liter of water per day (approximates the 90^{th} percentile) and a 70 kg adult consumes 2 liters of water (approximates 80^{th} percentile) per day. That policy has been reconfirmed in the November 2, 1999 policy document, "Estimating the Drinking Water Component of a Dietary Exposure Assessment."

The values applied in LifeLine™ reflect this EPA policy. Values between the 10 kg infant and 70 kg adult are approximated, consistent with trends of consumption for children in those age groups. These values are imprecise estimates designed primarily to fit the EPA policy. Table 4-6 give the values used in the model.

Table 4-6. Daily Residential Tapwater Intakes		
Bin	**Population**	**Seasons (All Same)**
1	<1, nursing	1000 ml (1 liter)
2	<1, non-nursing	1000 ml (1 liter)
3	1-4	1500 ml (1.5 liter)
4	5-12	1500 ml (1.5 liters)
5	13-24	2000 ml (2 liters)
6	25 >	2000 ml (2 liter)

4.4.2 WATER CONSUMPTION FOR THE 1994-1996, 1998 DATA

The CSFII obtains information both on water directly consumed (drunk as water) and as an ingredient listed in the recipes for nearly 1400 foods. All water in these recipes is coded as "municipal water." The corresponding EPA translation files have explicitly excluded water (EPA uses a broader array of water classifications), and deferred the assessment of water intake to a separate exercise (ongoing).

As noted above, fixed values for daily ingestion of domestic water were assigned on the basis of age to each record in the 1989-1991 CSFII. For the 1994-1996, 1998 CSFII, the Development Team has augmented this approach with alternative files that use information collected in CSFII to determine intake of domestic water.

Because the CSFII approach to recording consumption of "municipal water" does not distinguish between domestic water and other water supplies (e.g., water at a commercial bottling plant), and because many CSFII recipes that include "municipal water" are unlikely to be prepared at home on a regular basis, judgment is needed in deciding when to use (or not use) municipal water reported in CSFII as a source of exposure to domestic water. Two options have been implemented for the 1994-1996, 1998 CSFII in Version 2.0 of LifeLine™, in addition to the age-based assignment that is provided for both the 1989-1991 and the 1994-1996, 1998 data. Thus, the user has three options for water consumption when the more recent CSFII data are used for dietary assessment:

1. Fixed consumption (as used for the 1989-1991 data)
2. All municipal water consumed at home.
3. Municipal water likely to reflect domestic source.

In each of the latter two cases, water consumption is derived from the same CSFII record used to determine the consumption of foods. For each CSFII record, water consumption is calculated as the sum of:

- The fraction of direct water ingestion reported as coming from home and tapwater, and
- The "municipal water" contained in each home-prepared food, reflecting the amount of food consumed and the USDA recipe for that food.

The options differ with regard to how much of the "municipal water" coded in the record as a component of food is included in the estimate of domestic water consumption.

4.4.2.1 Direct Consumption of Domestic Water

The CSFII (Record Type 25) presents three data fields about direct water ingestion for each person and survey day:

- amount of water
- water from home
- away from home water

The first two of these underlie the calculation of domestic tapwater ingestion for each person and survey day. The third addresses the *kinds* (e.g., tap, bottled) of water consumed away from home, and is not considered further.

4.4.2.1.1 Amount of Water

This is a straightforward report of the volume of water ingested (fluid ounces) on each survey day by the respondent. It is the basis of further calculations of direct ingestion of tapwater. It includes "tap water or any bottled water that is not carbonated, with nothing added to it."

4.4.2.1.2 Water from Home

This categorical field addresses the proportion of water in the preceding variable that came from home:

1 = All
2 = Most
3 = Some
4 = None
8 = Don't know
9 = Not ascertained

4.4.2.1.3 Direct Ingestion Decision Rules

In order to obtain a quantitative estimate of direct ingestion of domestic tapwater, two estimations are required. First, the categories must be assigned fractional values, so as to quantitatively adjust the total amount of water consumed. This assignment is only unambiguous for categories 1 (all) and 4 (none). Second, a decision must be made as to whether or not commercially bottled water accounts for any fraction of this water ingestion, and if so, what fraction.

For Version 2.0 of LifeLine™, the following rules were applied:

1. All water was assessed as tapwater (no bottled water)
2. The assignment of fractional quantities to category codes was as follows:

Code	Description	Fraction Assigned
1	All	1.0
2	Most	0.75
3	Some	0.5
4	None	0.0
8	Don't know	0.5
9	Not ascertained	0.5

4.4.2.2 Water from Food: All Municipal Water Consumed at Home

There are 1,383 recipes for CSFII foods that list "municipal water" as an ingredient. These include both foods that are often prepared at home using tapwater (e.g., various brands of "infant formula, prepared from powder") and foods that are much less likely to be prepared using domestic tapwater (e.g., "Salisbury steak dinner, NFS (frozen meal)", or "Ice cream, soft serve, chocolate"). Some recipes specify the use of dry milk and municipal water, in cases where it seems likely that whole milk might be used instead (e.g., many of the ice cream recipes have this property). Others are clearly reconstituted with water, but the reconstitution could take place either at home or elsewhere. Orange juice, for example, is often reconstituted commercially, and then sold "ready to drink."

Moreover, each of these foods may come from a variety of sources, with implications for the inclusion of domestic tapwater. Food obtained at a restaurant, for example, may be significantly less likely to contain *domestic* tapwater than the same item prepared at home.

This analytical option only excludes those foods that are likely to be obtained "ready to eat," because of their source, and are therefore unlikely to involve the use of domestic tapwater.[46] Thus, foods reported as coming from the following sources (Record Type 30, Variable FOODSRCE, positions 100-101) are not used in the calculation of domestic water consumption:

2 = Restaurant with table service
3 = Fast food place, pizza place
4 = Bar, tavern, lounge
5 = School cafeteria
6 = Other cafeteria
7 = Vending machine

[46] This approach has the effect of excluding exposure to residues in water outside of the home (e.g., in a restaurant). While this assumption is not entirely plausible, it is more appropriate than assigning the residues found in a person's domestic water supply to every source of water encountered by that person.

8 = Child care center, family day care home, adult day care

16 = Residential dining facility

Any food from another source *is* used in the estimation of domestic water ingestion associated with that record.

4.4.2.3 Water from Food: Municipal Water Likely to Reflect Household Source

In this option, the calculation of daily water consumption from the CSFII record excludes a number of foods, based upon the professional judgment of the Development Team that they are unlikely to be prepared using domestic tapwater. Clearly, this approach always yields estimates of domestic water consumption that are less than or equal to those corresponding to the preceding option.

4.4.2.3.1 Foods that are Never Included

Any food with a name/description containing the following terms has been excluded from the calculation:

- Frozen (meal, entrée, dinner, dessert)
- NS as to made with milk or water
- NS as to dilution
- Canned, unless specifically noted as made with water

4.4.2.3.2 Foods that are Included

Once the food terms listed above have been excluded, the presence of one or more of the following terms in a food's description is used as the basis of including it in the list of foods used to estimate domestic water consumption:

- "Reconstituted," "diluted," "water added," "made with water," "salt removed in water," "...and water"
- "Prepared from liquid concentrate," "prepared from powder," "from boxed mix," "made from powder," "made from concentrate"
- "Evaporated milk," "condensed milk"
- "Rice," "paella"

- "Soup," "Chowder" (Unless noted as "ready to serve" or "NS with regard to dilution")
- "Made from dry eggs"
- "Beans, dry," "dry beans," or "chickpeas, dry"
- "Macaroni," "noodles," or "pasta," "ravioli," "gnocchi," "spaghetti," "manicotti," "stuffed shells," "tortellini"
- "Grits," "cereal, cooked," "cornmeal mush," "masa harina," "oatmeal, cooked," "cream of wheat," "whole wheat cereal cooked,"
- "Matzo ball"
- "Coffee," "tea"

4.4.2.4 Calculation of Water from Food (liters = kg)

For each dietary record (a given person on a given day), the consumption of water from food is calculated as the sum of the product for each food of its ingested mass and water content. Thus, if a person consumed the following three items in a day:

Code	Description	Amount	Water Fraction
92103000	Coffee, made from powdered instant, regular	430 grams (milliliter)	84.6%
56202980	Oatmeal, cooked, regular, NS as to fat added in cooking	740 grams	66.7%
58145115	Macaroni or noodles with cheese, from boxed mix with already prepared cheese sauce	1430 grams	4.9%

She would consume 927 milliliters of water as an ingredient in food. This amount would be added to the water reported as directly ingested.

4.5 Frequency of Selection

Human activities vary from individual to individual and over time for any single individual. Historically it has not been feasible to survey large numbers of individuals

over long periods of time. Therefore, the activity patterns of an individual over multiple days must be modeled using data from short-term studies.

The data from short-term studies inherently confound variability due to differences between the surveyed individuals, and the individual's day-to-day variability that occurred on the specific days measured in the survey. Thus, for any two records of different individuals on different days, one cannot say what fraction of the observed differences reflects enduring differences between the individuals and what fraction reflects the day-to-day variability in the behavior of each individual (Buck et al, 1995).

In order to model the day-to-day variation in any individual, one needs to collect data over a long period for each individual. Future versions of LifeLine™ could take advantage of such predictions when they become available.

While an explicit model of day-to-day variability in an individual's behavior is beyond the scope of available resources, it is possible to generally characterize the uncertainty that comes from this lack of data. This characterization is achieved by bounding the plausible range of exposures that could occur to an individual if longitudinal data were available[47]. This bounding is performed by adopting assumptions that minimize and exaggerate day-to-day variation. The two assumptions are as follows:

- Attribute all of the variability between records of similar individuals (as defined above) to intra-individual variability; or
- Attribute all of the variability to inter-individual variability.

The first option can be implemented by simply choosing a new record from the bin of similar individuals every day (the random option). The second option (the fixed option) can

[47] Note this approach does not fully bound the theoretical range of possible doses. If an individual's intake or behavior is negatively correlated with the prior day's intake or behavior it is possible that the short-term measures may be slightly higher and the long-term doses slightly lower that this version of LifeLine™ predicts.

be implemented by selecting a single record and holding that record constant (or fixed) day after day, until the individual changes bin (i.e., at a season/region (SR[48]) change).

To the extent that these extreme alternatives yield similar results with regard to exposure and risk, one can ignore the lack of data on day-to-day variation in activity pattern or diet. Alternately, one can obtain a measure of the range of possible outcomes for the analysis that might be observed given better data on individual consistency. To the extent that large variations between the two options are consistently observed in a range of assessments, one can argue that additional resources should be devoted to the collection of longitudinal data.

LifeLine™ allows the user to implementing the two options for sampling both activity and dietary records. Under the fixed residential option, LifeLine™ retrieves a single weekday record and a single weekend day record from the appropriate bin in NHAPS for an SR period. The individual is assumed to follow those records for each weekday and each weekend day for the SR period. Similarly, under the fixed dietary option, the system pulls a single CSFII record from the appropriate bin, and holds diet constant for the season. Under the random options for either dietary or residential exposures, LifeLine™ selects, with replacement, a different record for each day (for activity records, sampling always tracks whether the sampled day is for a weekday or weekend).

4.6 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.

US EPA. 1999. Overview of Issues Related to the Standard Operating Procedures for Residential Exposure Assessment, Presented to: FIFRA Scientific Advisory Panel For the meeting on September 21, 1999, United States Environmental Protection Agency Office of Pesticide Programs

[48] See Section 3A. above for a description of the SR.

US EPA. (1992) *National Home and Garden Pesticide use Survey*. Prepared by the Research Triangle Institute for the Office of Pesticides and Toxic Substances, Biological and Economic Analysis Branch.

CHAPTER 5. PESTICIDE RESIDUES[49]

Characterizing an exposure to an AI requires data on the concentration of a pesticide residue in the exposure media (food, air, water, and surfaces) and the behavior that brings the individual into contact with the residue. The previous chapter addresses behaviors. This chapter addresses the ways in which LifeLine™ models the presence of pesticide residues in food, in residential environments, and in tapwater.

5.1 Residues in Foods

This chapter will use a number of terms that have specific meanings in the description of pesticide residues in food (see text box).

In order to assess dietary exposures to residues of an AI, it is necessary to have data both on behavior (what foods people eat, and in what combinations at any given time), as described in Chapter 4, and on the residues in those foods. For most AIs of interest, however, representative data on the actual AI residues in foods as consumed are not available. This is obviously true for the prospective evaluation of new AIs being considered for registration but not yet actually in use in agriculture. Still, to anticipate the potential exposure from the use of a pesticide on agricultural crops (in the field), the anticipated residues on the foods (at the plate) must be modeled.

[49] LifeLine™ Version 2.0 has focused on pesticide active ingredients. The model, however, can be applied to any substance (active ingredients, inerts, food additives, or naturally occurring compounds) where the concentrations of the substance can be defined in Commodities or foods.

Definition of Key Terms

Food: Items eaten by humans; comprised of simple, raw commodities or complex combinations of multiple commodities, raw or processed. These are the items described as eaten by people in the food surveys utilized in the software.

Food Consumption: The amount of food eaten during one or more eating occasions by a single person. This can be viewed over many different time-periods ranging from a single eating occasion to an annualized (or lifetime) average.

Commodities: Agricultural items as whole crop items or subparts of a crop item. This is referred to as a "Raw Agricultural Commodity" or RAC for EPA Tolerance definitions.

Food Residues: The residue(s) present in foods (as eaten) that occur as the result of residues on the ingredients of the food.

Active Ingredient (AI): An active ingredient is a chemical and all metabolites of that chemical included in the residue analysis techniques employed in residue assessments. The exact definition for an active ingredient is provided in the definition of Tolerances as part of the EPA rulemaking process. Other definitions of the moieties included in an AI definition may exist, as related to the techniques employed to measure a residue of that AI, or in the legal definitions of that AI under different regulatory authorities.

In modeling the anticipated residues in foods as eaten, we must consider two key concepts. First, the foods as eaten may be complex foods comprised of many ingredients, each of which may represent only a fraction of the original agricultural commodity. For example, cookies may contain eggs, milk, wheat flour, corn oil, and chocolate (which itself is a combination of milk, sugar and cocoa). Thus, we must have accurate recipe files and information about the potential crop sources of refined, blended commodities such as sugar. Second, we must consider the processing steps imposed on a crop as it travels from the field to the plate. Even a simple, raw commodity such as an apple could be washed, waxed, and stored for a time before we purchase and eat it. The vast majority

of our food has more vigorous technologies applied—drying, bleaching, heating, clarifying, etc. Each processing technology has the opportunity to decrease or increase the concentration of the AI in all or a part of the commodity.

LifeLine™ separates the consideration of residues in foods as eaten from the consideration of dietary behavior. The calculation to predict residues in or on foods at the plate resulting from residues in the starting-agricultural commodities is addressed in the Food Residue Translator. The data about food consumption patterns is used in the LifeLine™ model—the main body of this program where the contributing information is combined and exposure/risk calculations are performed. The Food Residue Translator utilizes the information about AI use, residue information, and processing factors to calculate the distribution of residues possible in each food item as eaten. This provides an independent evaluation of the sources of variability and uncertainty in these two components of dietary exposure. The assessor is able to see whether variability in the analysis is primarily determined by residue or dietary variability. In addition, by viewing the distributions of residues predicted to exist in each food item as eaten, the assessor can consider if such predictions are plausible.

If the user has credible data on residue levels in foods as consumed, of course, these can be used directly in the analysis of dietary exposure, following appropriate formatting. Information about residue levels, processing effects, and pesticide use that has been formatted as ASCII files for use in software such as DEEM™ or Calendex™[50], can be used in this version of LifeLine™. This program will convert the information into the appropriate form for direct incorporation into the LifeLine™ analysis.

In summary, the input data for this analysis consist of data on the residues in food ingredients, which can either be entered for the specific form of the ingredient used in the food or predicted from data on raw ingredients and the effects of processing. A set of

[50] Propietary and trade secret software distributed by Novigen Sciences, Inc. Files applicable for use in LifeLine™ are not trade secret formats, rather are ASCII file formats, used by DEEM™ and Calendex™ programs for data storage.

recipe or translation files[51] is central to this analysis. Recipe files, which have hitherto been held top-secret, have been made visible via the LineLine project. Translation files have now been produced by a joint USDA / EPA effort, and are publicly available. Once the input data have been combined with the recipe or translation data, specific predictions of residues in foods (matched to the food coding from consumption surveys) are available.

5.1.1 INPUT DATA: CROP GROUPS, COMMODITIES, AND FOOD FORMS

LifeLine™ has adopted the EPA system for naming and categorizing the commodities on which pesticides may be used. This system reflects the opportunities for pesticide application to commodities or concentration change as the commodity travels from the field to the dinner plate. It also categorizes agricultural crops into botanically similar groups, within which surrogate data may be reasonably applied. The *Food Residue Translator* system reflects this hierarchical arrangement of residue data and related data.

5.1.1.1 Levels in the Hierarchy

There are three levels to the classification of residues and related data:

- A *Crop Group* organizes a set of related crops, such that data from one of these may be used to predict comparable data for the others. Examples include Small Fruits and Berries, Citrus Fruits, Pome Fruits (apples and pears), and Cereal Grains. The organization of a Crop Group reflects an inseparable blend of botany and agricultural practice (botanically similar crops are likely to have similar chemical retention properties).

- A particular crop is identified as a *Commodity (Raw Agricultural Commodity or RAC for 1989-1991, although the term is also used as a super category for food forms that are not in fact raw)*. The distinction between Commodities is based more on agricultural practice than on botany. For example, apples, dried apples, apple juice,

[51] Recipe files is the traditional term, now being replaced in official usage with the term "translation files". In the rest of this chapter, we will use recipe files to refer to the 1989-1991 CSFII translation process, and translation files to refer to that for the 1994-1996, 1998 CSFII.

and apple juice concentrate are distinct Commodities, rather than forms of the same Commodity.

- Food forms reflect the fact that substantial amounts of processing may intervene between harvest and food as consumed. Moreover, the individual constituent items of any food may have radically different amounts of processing. For example, the recipe for carrot cake includes 11 constituents, representing five designated as "cooked," four labeled "baked," and one each raw and raw-dried. Depending upon the extent to which a residue may be concentrated by drying or destroyed (if heat-labile) by cooking or baking, the predicted residues in the carrot cake may vary significantly from what would predict on the basis of residues in Commodities per se.

5.1.1.2 Changes in the 1994-1996, 1998 CSFII and Translation Files, Relative to the 1989-1991 CSFII and Recipe Files

While the basic hierarchy for classifying residues and related data has the same structure for the new data as for the old, there are important changes in the translation files, relative to historic practice with recipe files (see box). At a minor level, the assignment of some Commodities to Crop Groups has been changed in the translation files, relative to that used previously for the recipe files. More importantly, the meaning and usage of the Food Form concept has changed dramatically.

The Redefinition of Food Form in the USDA / EPA Translation Files

With the introduction of the USDA / US EPA translation files for the 1994-1996, 1998 CSFII data, the interpretation of Food Form became considerably more complicated, and the use of Food Form in the prediction of residues in foods from residues in Commodities became far more difficult. Before using the *Food Residue Translator* to make such a prediction, the user should be familiar with the change from historic practice that is embodied in the new translation files. As noted above, the form in which a Commodity occurs in a food may have a dramatic influence on the residues that it contributes to that food. For example, if the heat treatment associated with commercial canning destroys a residue, one would expect to see lower residues in a food made with a canned form of a Commodity than in a similar food made with the fresh form of that Commodity.

Historic Practice: Historic recipe files have facilitated the inclusion of such considerations in the analysis. Food Forms were assigned independently to each ingredient (RAC) in the recipe, such that the influence of processing specific to different food forms could be appropriately addressed in estimating residues in a food.

Practice in the translation Files: The USDA/EPA translation Files depart radically from this historic practice. Rather than an ingredient-specific Food Form, each recipe in the files contains a Cooking Status - Food Form - Cooking Method (CSFFCM) that is assigned to the *food*, and then imputed to every ingredient in the food. Thus, it is impossible to combine different Food Forms within a food. The problems for predicting residues in many foods are apparent. For example, in a chicken salad, the chicken is generally cooked while the greens are generally raw. These kinds of Food Form relationships cannot easily be represented in the new translation files.

In order to (at least partially) address this issue, the number of CSFFCM categories has increased dramatically. While there are a total of 24 Food Forms in the recipe files for the 1989-1991 CSFII, there are 512 possible CSFFCM combinations for the translation files, of which 73 are actually used.

The fact that every ingredient in a food *must* have the same CSFFCM calls for great care on the part of the system user in assigning different residues to differing forms of the same Commodity. It is possible to assign different residues, processing, or use factors to CSFFCM categories when dealing with the 1994-1996, 1998 CSFII data, just as it is for Food Forms when applying the historic recipe files to the 1989-1991 consumption data. The application of these data, however, may be very different.

Consider the case where different residue values are applied to the two forms of fresh basil included in the translation files:

cooked/not specified/not specified,
uncooked/fresh/N/A

Unlike the case of the earlier recipe files, with the translation files, the former data will only be applied to foods that have been cooked; the latter data will only be applied when the whole food is judged fresh. In the specific case above, the cooked fresh basil shows up

in only one food (581471000: Pasta with pesto sauce), and the uncooked fresh basil also shows up in only one food (813020700: Pesto sauce). As pesto sauce is generally applied to pasta *after* the pasta is cooked, and subject to incidental heating only, one would not generally expect different residue contributions in these two cases.

In more general terms, the problem exists when either cooked or uncooked forms of the same Commodity may appear in cooked foods. The user should familiarize herself with the actual recipes in the translation files prior to entering data for different CSFFCM categories of a Commodity.

5.1.1.3 Data Types

Ideally, each user could supply data in terms of a residue distribution for each food form of a particular Commodity (taking into account how these are represented in the recipe or translation files). In most cases, however, such a complete set of residue data would be lacking. The *Food Residue Translator* supports the inclusion of two additional data sets that allow the user to predict residues in the ingredients of each food, a *use Probability Factor* and a set of *Processing Factors*.

5.1.1.3.1 Residue Data

In order to predict residues in foods, the user must be able to supply residue data at some level of the hierarchy (specific Food Form of the Commodity, data for all forms of the Commodity, or generic data for the entire Crop Group).

5.1.1.3.1.1 Data Sources

Absent measured data, it is possible to use Tolerances, which represent regulatory maximum concentrations that if exceeded indicate a use in excess of label requirements. Tolerances are legal limits, formally assigned by EPA for each use of each AI. Several field trials are performed on the commodity to determine the maximum residue that could result from label use of the AI. The Tolerance is set just above that maximum. Thus, it is a number higher than any maximum ever measured from such trials. Using Tolerances

will thus certainly lead to an overestimate of potential food residues. Tolerances are almost equivalent to the Maximum Residue Levels set for such crops under the World Health Organization's CODEX Alimentarius system and by other governments.

Another source of publicly available data are the Pesticide Data Program (PDP) files distributed (with a search utility) by the US Department of Agriculture. A copy of this data set has been included on the LifeLine™ CD-ROM. This file contains collections of residues recorded from a nationwide crop-monitoring program. In many cases, the PDP data will represent "non-detects." These data will generally have to be represented by "proxy" concentrations that are reflective of the detection limits applicable to each set of samples (a conventional proxy value is one-half of the detection limit for the sample). In this and other issues, appropriate selection and use of these data require considerable judgment on the part of the user. Such deliberations and decisions should be carefully documented.

5.1.1.3.2 Use Probability Factors

In many cases, residue data are representative of the residues that one might encounter in Commodities that have been treated with the AI of interest. Such data will not reflect the fact that for some AIs, only a fraction (sometimes a very small fraction) of the crop in any season will have been treated with the AI.

By specifying a *use Probability Factor* for a particular Food Form, Commodity, or Crop Group, the user can directly account for the fact that the expected residues in some instances will truly be zero.[52]

Assessments generated by EPA will indicate the use probability factors that were employed (if any) in the public docket, along with any residue data used in the analyses.

[52] Accordingly, the LifeLine™ system maintains a logical distinction between "non-detects" (samples where the AI might be expected to be found, but are not detected at some analytical limit of detection) and "true zeroes" (instances where there is no *a priori* expectation of a residue).

5.1.1.3.3 Processing Factors

As noted above, the processing of a Commodity may have a major influence on pesticide residues. LifeLine™ addresses the following distinct Food Forms in its recipes:

Table 5-1. Food Forms for 1989-1991 CSFII	
11	Raw
12	Cooked
13	Baked
14	Boiled
15	Fried
16	Pasteurized
18	Raw: Dried
19	Cooked: Dried
32	Canned: Cooked
33	Canned: Baked
34	Canned: Boiled
35	Canned: Fried
38	Canned: Raw/Dried
39	Canned: Dried/Cooked
41	Frozen: Raw
42	Frozen: Cooked
43	Frozen: Baked
44	Frozen: Boiled
45	Frozen: Fried
48	Frozen: Dried-Raw
49	Frozen: Dried-Cooked
51	Smoked/Cured/Salted/Raw
52	Smoked/Cured/Salted/Cooked
59	Smoked/Cured/Dried

Table 5-2. Cooking Status, Food Form, Cooking Method for 1994-1996, 1998 CSFII		
Status:	**Form:**	**Method:**
uncooked	fresh	N/A
uncooked	fresh	boiled
uncooked	frozen	N/A
uncooked	dried	N/A
uncooked	cured, pickled, smoked, salted	N
uncooked	other	N/A
uncooked	Not specified	N/A
cooked	fresh	N/A
cooked	fresh	baked
cooked	fresh	boiled
cooked	fresh	fried
cooked	fresh	fried or baked
cooked	fresh	boiled or baked
cooked	fresh	not specified
cooked	fresh	NS as to further cooking
cooked	frozen	N/A
cooked	frozen	baked
cooked	frozen	boiled
cooked	frozen	fried
cooked	frozen	fried or baked
cooked	frozen	not specified
cooked	frozen	NS as to further cooking
cooked	dried	N/A
cooked	dried	baked
cooked	dried	boiled
cooked	dried	fried
cooked	dried	fried or baked
cooked	dried	not specified
cooked	dried	NS as to further cooking
cooked	canned	N/A
cooked	canned	baked
cooked	canned	boiled
cooked	canned	boiled or baked
cooked	canned	not specified
cooked	canned	NS as to further cooking
cooked	cured, pickled, smoked, salted	N/A
cooked	cured, pickled, smoked, salted	baked
cooked	cured, pickled, smoked, salted	boiled
cooked	cured, pickled, smoked, salted	fried
cooked	cured, pickled, smoked, salted	boiled

Table 5-2. (Cont.) Cooking Status, Food Form, Cooking Method for 1994-1996, 1998 CSFII		
Status:	**Form:**	**Method:**
cooked	cured, pickled, smoked, salted	not specified
cooked	cured, pickled, smoked, salted	NS as to further cooking
cooked	other	N/A
cooked	other	baked
cooked	other	boiled
cooked	other	fried
cooked	other	fried or baked
cooked	other	not specified
cooked	other	NS as to further cooking
cooked	not specified	N/A
cooked	not specified	baked
cooked	not specified	boiled
cooked	not specified	fried
cooked	not specified	fried or baked
cooked	not specified	boiled or baked
cooked	not specified	not specified
cooked	not specified	NS as to further cooking
frozen meal	frozen	baked
salad	fresh	N/A
salad	fresh	not specified
salad	other	not specified
salad	not specified	N/A
salad	not specified	boiled
salad	not specified	not specified
sandwich	not specified	N/A
sandwich	not specified	fried
sandwich	not specified	not specified
not specified	frozen	N/A
not specified	dried	N/A
not specified	other	N/A
not specified	not specified	N/A
not specified	not specified	not specified
unknown		

The system supports user entry of up to six separate processing factors for any food form. The default labels (which the user may change) are:

- Dehydration;
- Washing;
- Heating;

- Refining;
- Storage; and
- Other.

For example, the user may replace default labels here with processing technologies such as juicing or bleaching. The EPA has published a listing of default values representative of some processing technologies. These are based on EPA's experience with many residue studies employing such technologies. Those are provided in LifeLine™ (FACTORS.DBF) and can be downloaded into the "Other" column.

From the individual processing factors for each process, the system computes a net processing factor. For example, if the user specified the following processing factors for Apples / Canned:Baked:

Table 5-3. Example Food Factors Values

Factor	Value
Dehydration	1.25
Washing	0.75
Heating	0.8
Refining	
Storage	0.9
Other	

The system would generate a *net* processing factor of 0.675. Note that if a processing factor has not been specified, there is no assumed effect (default value of 1.0).

5.1.1.4 Specifying Data

The LifeLine™ Food Residue Translator supports either manual entry of data or file import of dBase files and ASCII files containing data previously prepared for use in DEEM™ and Calendex™ software. Obviously, where there are large amounts of data, file import is more efficient (and less error-prone) than manual entry.

5.1.1.4.1 Manual Data Entry

The main display of the system consists of three spreadsheets for data entry. The user may enter residue values (as mg of AI per kg of commodity), use probability factors (as decimal fractions), or processing factors (as decimal fractions representing the percent of the residue remaining after processing), depending upon which of the three spreadsheets is selected. The format of each spreadsheet is very similar, reflecting the fact that all data can be entered at any level of the hierarchy.

Each spreadsheet opens at the Crop Group level, and allows the user to expand a Crop Group to its constituent Commodities or a Commodity to its constituent Food Forms. If the user specifies data for an entire Crop Group, they will be applied to all members of that group for which specific data have not been supplied. For example, if the user specifies residues for "Citrus Fruits," these would apply to all Commodities (RACs) from grapefruit, lemon, lime, orange, and tangerines.[53]

Alternatively, the Crop Group display can be expanded to show its constituent commodities, so that data can be entered for a specific commodity (e.g., Orange Peel). In a similar fashion, an individual Commodity display can be expanded to show all constituent Food Forms. In the above example, selecting the row for Orange Peel and clicking on the expansion button (on the toolbar) will display the five food forms for this commodity utilized in the 1989-1991 recipe Files (canned:cooked, raw:dried, raw, cooked, baked).

For any Crop Group, Commodity, or Food Form, the user can simply type in a residue value into the first (or next empty) cell in the corresponding row of the residue spreadsheet. The entries must be consistent with the metrics specified for each field (ppm for residue concentration values, decimal fractions for the use Probability and Processing Factors. The spreadsheet allows the entry of hundreds of residue values in any row. Alternately, the user may enter a single use Probability Factor or up to six Processing Factors per row in the corresponding spreadsheets.

If the user has not supplied values for the probability of use, the system assumes that the entire commodity (100%) in the food supply has been treated. Also, if the user enters one or more zeroes (as opposed to non-zero proxy concentrations) as residues for a Crop Group, commodity or food form, any information supplied for that Crop Group, commodity or food form on probability of use is ignored (the user is warned when the first zero is entered). This is because a zero entered as a residue value implies that the non-treated commodities are already represented in the residue database and further modeling from use probability statistics is unnecessary.

Note that manual data entry can be aided with typical "Copy and Paste" functions, and "Drag and Paste" actions as in Microsoft™ spreadsheets.

5.1.1.4.2 Data File Formats

If an AI is used in a large variety of Commodities (and these cannot be addressed at the Crop Group level), if different values apply to various Food Forms, or if any commodities have a large number of residues, manual data entry will be inconvenient. This is particularly true if the data have already been stored electronically. Accordingly, the system supports data entry by means of an electronic (dBase format) file.

Data for diverse databases must be merged into a common dBase file before incorporation into the Food Residue Translator data screen. This can be accomplished in Excel and directions for such data manipulation are given in the User's Manual.

A single file is used to import data on residues, probability of use, and processing factors. Although the information captured is primarily numeric, it is supplied as text strings. The fields for this database are given in Table 5-4.

[53] EPA Commodity definitions include several forms of each of these fruits.

Table 5-4. Structure of Residue, Use, and Processing Input File				
Field	**Type**	**Length**	**Values**	**Notes**
Cropgroup	Character	2	Code	Mandatory
Raccode	Character	3/6*	Code	
Foodform	Character	2/3*	Code	
Coverage	Character	1	Code	Season code (or annual)[54]
Residue	Character	10	ppm (mg/kg)	
Pfactor1	Character	6	unitless	
Pfactor2	Character	6	unitless	
Pfactor3	Character	6	unitless	
Pfactor4	Character	6	unitless	
Pfactor5	Character	6	unitless	
Pfactor6	Character	6	unitless	
Upfactor	Character	6	0-1 (unitless)	
* Former value for 1989-1991; latter for 1994-1996, 1998				

The first three fields require the entry of numeric codes:

- Codes for Crop Groups can be read off of the system spreadsheets, and are also found in the system database *RTCGROUP.DBF*,
- Codes for Commodities are in *RTRAC.DBF*, and
- Codes for Food Forms are in *RTFFORMS.DBF*.

It should be borne in mind that while codes for Crop Groups and Commodities are unique, those for Food Forms are not (i.e., the same Food Form could apply to more than one Commodity). Each record in this database must have values for Crop Group, Food Form Data must have both Crop Group and Commodity.

[54] 0= annual, 1 = spring, 2 = summer, 3 = fall, 4 = winter (see below)

Each record in the database should represent a single residue, use probability factor, or set of processing factors. Care should be taken in entering the latter two types of values, as only a single value is used for each Food Form, Commodity, or Crop Group addressed. In the case of multiple entries of such records, only the last item in the file is retained.

To specify residues for a particular Food Form of a particular Commodity within a Crop Group, the user would enter codes for:

- Crop Group code (Mandatory)
- Raccode (If left blank, residue is assigned to the Crop Group)
- Foodform code (If left blank, residue is assigned to the Commodity [or Crop Group if no Commodity is specified])
- Coverage code (Mandatory: Controls assignment to particular season or annual file)
- Residue (Mandatory, Units of ppm = mg/kg)

All other fields in these records are left blank.

To specify probability of use for a particular Commodity, the user enters codes for:

- Crop Group code (Mandatory)
- Raccode (If left blank, probability of use is assigned to the Crop Group)
- Coverage code (Mandatory: Controls assignment to particular season or annual file)
- Upfactor (Must be between 0 and 1)

All other fields in these records are left blank.

Note that Food Form is not generally of interest when assigning data on probability of pesticide use.

To specify processing factors for a particular Commodity / Food Form, the user enters codes for:

- Crop Group code (Mandatory)
- Raccode (Mandatory)
- Foodform code (If left blank, factors are assigned to the Commodity)

- Coverage code (Mandatory: Controls assignment to particular season or annual file)
- Pfactor1 (Any nonzero value is valid, blanks are ignored [treated as 1.0])
- Pfactor2 (Any nonzero value is valid, blanks are ignored [treated as 1.0])
- Pfactor3 (Any nonzero value is valid, blanks are ignored [treated as 1.0])
- Pfactor4 (Any nonzero value is valid, blanks are ignored [treated as 1.0])
- Pfactor5 (Any nonzero value is valid, blanks are ignored [treated as 1.0])
- Pfactor6 (Any nonzero value is valid, blanks are ignored [treated as 1.0])

All other fields in these records are left blank.

5.1.1.5 Data Integration

Both manual entry and import of a database file support the entry of data at multiple levels in the hierarchy. When both generic and specific data have been supplied, the specific data take precedence, and generic data are used when no specific data are available. For example, one might specify one set of residue data that would apply to the Commodity "raspberries," and a different set of data for the Crop Group "small fruits and berries." In this case, the Crop Group values would be applied to all members of the Crop Group *except* raspberries. The specific data entered for raspberries would take precedence over the generic data for the Crop Group.

Prior to generating estimates of residues for foods as consumed, the system integrates the three types of information supplied by the user. Thus, entered residues are combined with data on the probability of use, so as to generate a complete set of potential residues (including both nonzero and zero values), and processing factors are combined into a net factor that is then applied to scale residues up or down.

5.1.1.5.1 1) Adjusting Residue Distributions for Probability of Use

The assumption made in defining residue distributions for the system is that there are two types of "non-detect" events for pesticide residues that require different strategies. In some cases, the commodity in question was not treated with the particular AI, so that it is reasonable to assume a zero residue. In other cases, a pesticide containing the AI may have been applied, but the AI was not detected in subsequent analysis, in these cases, the user may choose to use a proxy concentration corresponding to the sensitivity of the analytical method should be used.

For any data set in which a probability of use (e.g., percent crop treated) has been specified along with residue data, the system adjusts the residue distribution, to reflect the value assigned to the probability of use. For example, if the user specified five (nonzero) residue values for a Commodity, and a value of 0.10 for probability of use on that Commodity, the system would create a final residue distribution for the Commodity consisting of the five nonzero values and 45 zeroes.

5.1.1.5.2 Modifying Residue Data with Net Processing Factors

While the system allows the user to enter as many as six separate processing factors, the important fact for the evaluation of residues in foods is the *net* processing factor that applies to the particular Food Form of the Commodity that is used in a food. Accordingly, the system cross-multiplies the processing factors to determine the net factor, as described above.

5.1.1.5.3 Propagation of Residue Data

As noted above, data can be entered at any level of a hierarchy of description: Crop Group, Commodity, or Food Form. Where data have been entered at the most specific level (Food Form of a Commodity), they are applied *only* to that specific food form. The most common instance would be processing factors.

Where data have been entered at a more general level (Commodity or Crop Group), they are assumed to apply to *all* specific instances, *unless* specific data have been provided for

that specific instance. Thus, for example, if the user specified residues (and probability of use) for Apples, these values would apply to all of the relevant Food Forms:

1989-1991	**1994-1996, 1998**
Apple	***Apple, fruit with peel***
• Frozen:Boiled	• cooked; fresh; fried
• Frozen:Cooked	• uncooked; fresh; N/A
• Canned:Boiled	• cooked; not specified; baked
• Cooked	• cooked;f fresh; baked
• Canned:Cooked	• uncooked; cured, pickled, smoked, salted; N/A
• Fried	• salad; fresh; N/A
• Baked	
• Raw	
• Canned:Baked	

Unless different values had been specified for one or more of these specific food forms.

Similarly, if the user specified residues for ***Pome Fruits***, they would be applied to:

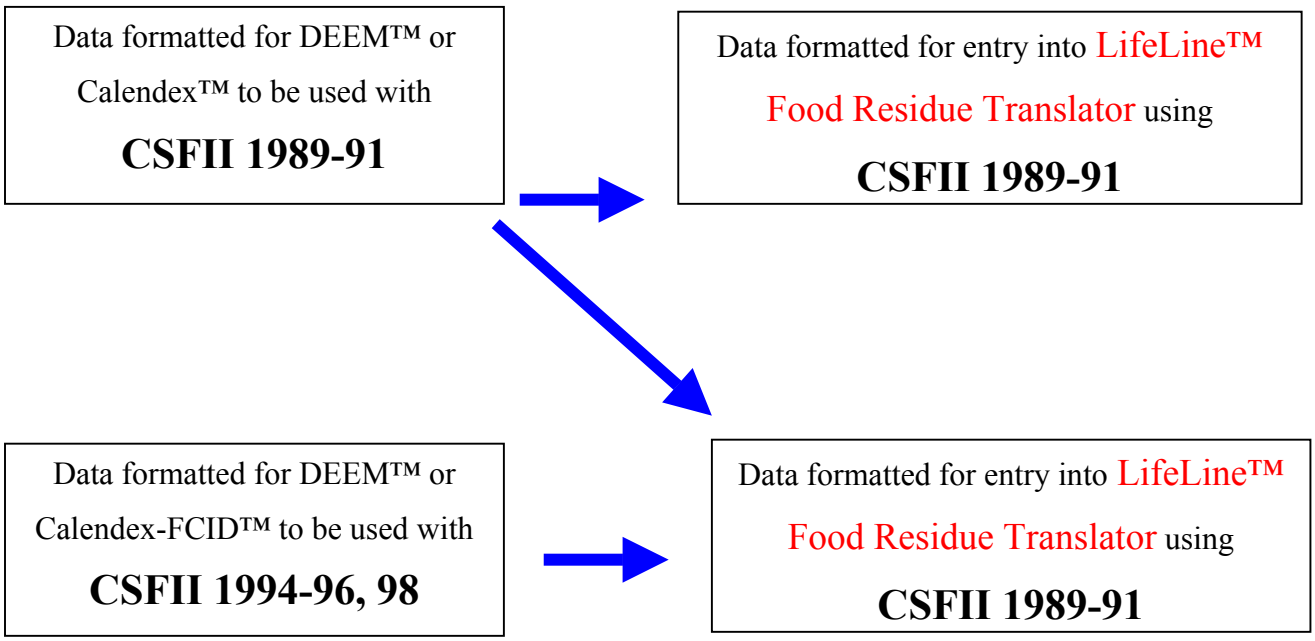
1989-1991	**1994-1996, 1998**
• Apples-Dried	• Apple, peeled fruit
• Apples-Juice/Cider	• Apple, peeled fruit- babyfood
• Pears	• Apple, dried
• Pears-Dried	• Apple, dried - babyfood
• Apples-Juice-Concentrate	• Apple, juice
• Pears-Nectar	• Apple, juice - babyfood
• Apples	• Apple, sauce
	• Apple, sauce - babyfood
	• Crabapple
	• Loquat
	• Quince
	• Apple, fruit with peel

Again, any value entered for a *specific* instance (e.g., Pears-Nectar) would take precedence over the generic Pome Fruit value.

5.1.2 CONVERSIONS OF DATA FORMATTED FOR DEEM™ OR CALENDEX™ SOFTWARE TO LIFELINE™ FORMATS

The formatting and category strategies used in other dietary exposure models for crops, crop groups, food forms processing factors and other modifying parameters are different than those described above. Therefore, in order to use the data files formatted for use in those models, LifeLine™ must convert information from those file fields into a format appropriate for the LifeLine ™ model. Because there are multiple versions of DEEM™ and Calendex™, multiple conversion strategies were employed. Those have been developed and are available in LifeLine™ Version 2.0 and its subsequent versions. Those strategies are called "Bridges" for conversational convenience. A summary of those Bridges is presented in Figure 5-1.

Figure 5-1



Instructions for importing the data from these ASCII formats into LifeLine™ is presented in the Users Manual. The process is quite simple to do, but the underlying Bridge structures should be understood by the user. The logic of those conversions is presented here, with complete documentation in the Appendices of this manual.

The general idea is to match fields in the ASCII files that correspond to similar information fields in the LifeLine™ data formats. Values representing crop group residues, or individual crop residues, processing factors or Percent Crop Treated values will be found in the ASCII files and applied to the appropriate LifeLine™ data field.

There are two inconsistencies that have made it impossible to create perfect conversions. There are two sets of dietary consumption surveys in use by both programs (CSFII 1989-1991 and CSFII 94-96,98). Each has its own set of food definitions and recipe files with processing factor logic. Secondly, the category logic for food forms, crop definitions and modifying factors are not exactly the same in these different models. The bridge employs reasonable strategies for linking similar fields, but reasonable people may prefer alternative linking rationale. To accommodate this, the exact linkage pathways are presented in the Appendices of this Manual to inform the user of the strategy employed in the three sets of Bridges. After the data are imported into the LifeLine™ Food Residue Translator, the user may amend the converted input file prior to use in the LifeLine Program calculations.

Also, after an importation of data using the Bridges, the Food Residue Translator will provide a listing of "Warnings". These are not errors, rather a notation to highlight places where the conversion was not exact, or where no conversion was possible. The user can consider these issues and apply corrections to the converted file as they think is appropriate.

Error warnings will also be listed separately if true errors occur in reading the import file or making the conversion. These are not conversion logic highlights. Error warnings

denote a malfunction in the process of making the conversion and should be resolved completely before continuing with the exposure and risk analysis.

LifeLine™ Version 1.0 and 1.1 used the 1989-91 CSFII Dietary Survey as the basis for its dietary exposure and risk calculations (Version 1.1 also included the 1994-96, 98 survey). That entire function, along with the recipe files appropriate for the 1989-91 survey[55] have been incorporated into LifeLine™ 2.0, providing a choice to use this survey or the 1994-96, 98 CSFII Dietary Survey. If the user starts with crop residue files (and associated processing and use information) on ASCII files created from the 1989-91 survey format, they can convert the information into a LifeLine™ file to be used with the 1989-91 CSFII data (Bridge 1) OR convert it into a LifeLine™ file to be used with the 1994-96, 98 CSFII data (Bridge 2).

BRIDGE 1: ASCII files using CSFII 89-91 food listings TO LifeLine files using CSFII 89-91 food listings

The files being imported are ASCII files formatted for use with the USDA CSFII 1989-91 survey. These file structures contain information about residues in or on crop groups, commodities or food forms, information about modifying factors such as processing factors, and information about percentage of pesticide use, listed with each crop group, commodity or food form field. These files would be consistent with those originating for use in DEEM™ or Calendex™ dietary software versions using CSFII 1989-91 dietary survey data in the exposure and risk analysis.

The conversion will result in a LifeLine™ file structure that is appropriate to use with the same USDA CSFII 1989-91 dietary survey for the exposure and risk analysis to be performed by the LifeLine™ program.

[55] All recipe files in LifeLine™ software can be viewed in the Knowledge Base files located on the original disc or with the electronically conveyed files for LifeLine™. All such data elements are visible in LifeLine™ software.

The complete conversion strategy is presented in the following Appendices to this manual.

Appendix A:

This appendix shows the conversion for each commodity listed in the imported database to the LifeLine™ codes.

Source 1 RACCode is the code used in the originating ASCII file (consistent with the descriptors in DEEM™ or Calendex™ software for a single raw agricultural commodity (RAC)

Source1Description is the name of this RAC as defined in the DEEM™ or Calendex™ software

LifeLineRACCODE is the code to which the Source 1 commodity is being converted for LifeLine™

LifeLineCrop Group is the Crop Group in LifeLine™ to which the Source 1 RAC is being converted

LifeLineDescription is the name of the commodity for that RAC Code as it appears in the LifeLine™ software.

Example from the complete listing in the Appendix file:

Source1RACCode	Source1Description	LifelineRACCode	LIfelineCropGroup	LifelineDescription
98	Acerola	97	1	KIWI FRUIT
316	Alcohol-distilled	316	1	DISTILLED ALCOHOL
248	Alfalfa	248	1	ALFALFA SPROUTS

Appendix B:

This Appendix shows the conversion of food forms as coded and described in the imported ASCII file to the code and description used in LifeLine™ format
The Source1FoodForms Code is the code used in the imported data format consistent with descriptions used in DEEM™ or Calendex™ versions that use the CSFII 1989-91 survey data.

The Source1FoodForms Description is the description used in the imported data format

The LifeLine10FoodFormsCode is the code used in this LifeLine™ software (originating from the LifeLine™ Version 1.0 format) into which the imported data is converted.
The LifeLine10FoodForms Description is the description of this food form as used in the LifeLine™ software for food forms in the CSFII 1989-91 dietary survey.

Example from complete listing in Appendix file:

Source1FoodForms.Code	Source1FoodForms.Description	Lifeline10FoodForms.Code	Lifeline10FoodForms.Description
11	Uncooked	11	Raw
60	Canned: Cured	52	Smoked/Cured/Salted/Cooked

BRIDGE 2: ASCII files using CSFII 89-91 food listings TO LifeLine files using CSFII 1994-96, 98 food listings

The files being imported are ASCII files formatted for use with the USDA CSFII 1989-91 survey. These file structures contain information about residues in or on crop groups, commodities or food forms, information about modifying factors such as processing factors, and information about percentage of pesticide use, listed with each crop group, commodity or food form field. These files would be consistent with those originating for use in DEEM™ or Calendex™ dietary software versions using CSFII 1989-91 dietary survey data in the exposure and risk analysis.

The conversion will result in a LifeLine™ file structure that is appropriate to use with the LifeLine™ exposure and risk analysis that utilizes the USDA CSFII 1994-96 dietary survey and its 1998 Supplement Survey for Children. It is important to understand that these CSFII surveys conducted by USDA are quite different in many ways. The foods listed as eaten by the participants are quite different in their descriptions for the different surveys. Also, food forms differ between the two surveys. These differences, along with the differences in the coding approaches among exposure models, yield a rather complex and imperfect bridging structure. The user may wish to amend the imported file as needed to conform with their own strategy for conversions.

The complete conversion strategy is presented in the following Appendices to this manual.

Appendix C:
Headers in Appendix C file:

BRIDGE 2: CONVERSION LOGIC FOR CROP GROUPS AND COMMODITIES

Source Data Elements			LifeLine Data Elements						
Source Crop Group Code	Source RAC Code	Source Description	Default CC (do not override)	Proc. Factor Linkage	Special Note	LL Crop Group	LL Commodity Code	2nd LL CC Equivalent	3rd LL C

The Source Data Elements show the crop group code and commodity code and description from the imported files. These are consistent with the DEEM™ and Calendex™ codes and descriptions used in versions compliant with the CSFII 1989-91 survey terms.

The LifeLine™ Data Elements show how these Source Data Elements were converted into codes and descriptions consistent with the LifeLine™ format compliant with the CSFII 1994-96, 98 survey terms. In many cases there are multiple codes in LifeLine™ data elements applicable to one term in the Source data. In such cases, a default code is chosen as the first term of preference. The processing factor linkage is presented which

corrects the residue values associated with this term appropriately. See the special notes, below, for further explanation. Special notes are included where direct conversions were not possible or where there was no conversion possible.

Column D: "Default CC (do not override)

In some cases more than one Same Commodity Code (CC) is equivalent to the same LifeLine™ Commodity Codes. If both Source Commodities have identical residue values, there is no problem. However, when the Source Commodities have different residue values, the one with the "d" remains fixed and is the one to be used in the equivalent LifeLine™ calculation.

Column E: Processing Factor Linkage

Some correction factors will be used with these Commodity Codes, and this defines the type. It is useful for tracking. These definitions are linked to the next column, "Special Note".

Column F: Special Note

This flags a calculation or note, relevant to the Processing Factor Linkage or other message. A list of these notations is provided below with their definitions.

Column G: LL Crop Group Code

Equivalence to Source Crop Group Code

Column H: Life Line Commodity Code

Equivalence to Source Commodity Code

Column I: 2nd LifeLine™ Commodity Code

If one Source Commodity Code equates to more than one LifeLine™ Commodity Code, this is the 2nd LifeLine™ Commodity Code to which it equates.

Column J-Column M Additional LifeLine™ Commodity Codes

If one Source Commodity Code equates to more than two LifeLine™ Commodity Codes, these are the LifeLine™ Commodity Codes to which they equate.

Special Notations in Column F

1. Concentrate * The equivalent LifeLine™ Commodity Code (LLCC) must be linked to a correction factor of 1/2 in order to correct from residue values found in concentrate to residues that would result in the reconstituted juice.

   [Source RAC residue values] [1/2] = residue values for equivalent LifeLine™ Commodity Code

2. Dried ** The equivalent LifeLine™ Commodity Code must be linked to a correction factor of 3.3 in order to correct from residue values found in the whole commodity to the residue values found in the dried LifeLine™ Commodity Code. (Raw Agricultural Commodity = RAC)

   [Source RAC residue values] [3.3] = residue values for equivalent "dried" LifeLine™ Commodity Code

3. Oil *** The equivalent LifeLine™ Commodity Code must be linked to a correction factor of 8 in order to correct from residue values found in the whole commodity to the residue values found in the oil LL CC.

   [Source RAC residue values] [8] = residue values for equivalent "oil" LL CC

4. Syrup **** The equivalent LifeLine™ Commodity Code must be linked to a correction factor of 3 in order to correct from residue values found in whole commodities to the residue values found in the syrup LifeLine™ Commodity Code.

[Source RAC residue values] [3] = residue values for equivalent "syrup" LifeLine™ Commodity Code

5. Note 1 "Pears and Pears, dried and Pears, juice" have been incorrectly coded by USDA into the "Stone Fruit" Crop Group. Check to be sure crop group residues are not misapplied to these commodities.

This note will appear in the *Warning List* when the bridge imports any values from these commodities into LifeLine™ Commodity Codes and also when a "Stone Fruit" Crop Group Residue is imported into LifeLine™ Commodity Codes.

6. Note 2 Source Code does not have an equivalent code in LifeLine™ Commodity Code for CSFII 94-96, 98.

This note should appear in the *Warning List* when one of these commodities from the source appears with a residue. There is no equivalent LifeLine™ commodity to apply those residues.

7. Note 4 "Raspberry and Raspberry, babyfood and Raspberry, juice and Raspberry juice babyfood" have been incorrectly coded by USDA into the Tree Nut Crop Group. Check to be sure crop group residues are not misapplied to these commodities.

This note will appear in the *Warning List* when the bridge imports any values from these commodities into LL CCs and also when a Tree Nut Crop Group residue is imported into LifeLine™ Commodity Codes.

Appendix D:

Example from complete file presented in Appendix D:

Source Data Elements		LifeLine Data Elements									
FoodFormCode	Description	LL Equivalent 1	LL Equivalent 2	LL Equiv 3	LL Equiv 4	LL Equiv 5	LL Equiv 6	LL Equiv 7	LL Equiv 8	LL Equiv 9	LL Equiv 10
15	Fried		213		214	283	284	293	294	893	
16	Pasteurized		no equivalent								

The Source Data Elements show the food form code and food form description from the imported files. These are consistent with the DEEM™ and Calendex™ codes and descriptions used in versions compliant with the CSFII 1989-91 survey terms.

The LifeLine™ Data Elements show how these Source Data Elements were converted into codes and descriptions consistent with the LifeLine™ format compliant with the CSFII 1994-96, 98 survey terms.

Note that more than one food form is applicable to a given source food form, and sometimes there is no equivalent food form in the 1994-96, 98 survey listing. In such cases, the user must adjust the resulting imported file in LifeLine's Food Residue Translator to accommodate their own considerations for these situations. If residue values are presented in the source data elements where there is no equivalent in the LifeLine™ version, those residue values will not factor into the risk assessment unless they are assigned elsewhere in the crop listing. Such situations will be brought to the user's attention in the *Warning Listing* that appears after conversions are completed in the LifeLine™ software data import function.

BRIDGE 3: ASCII files using CSFII 1994-96, 98 food listings TO LifeLine files using CSFII 1994-96, 98 food listings

The files being imported are ASCII files formatted for use with the USDA CSFII 1994-96 survey and its 1998 Supplemental Children's Survey. These file structures contain

information about residues in or on crop groups, commodities or food forms, information about modifying factors such as processing factors, and information about percentage of pesticide use, listed with each crop group, commodity or food form field. These files would be consistent with those originating for use in DEEM-FCID™ or Calendex-FCID™ dietary software versions using CSFII 1994-96,98 dietary survey data in the exposure and risk analysis.

The conversion will result in a LifeLine™ file structure that is appropriate to use with the same USDA CSFII 1994-96, 98 dietary survey for the exposure and risk analysis to be performed by the LifeLine™program. The Source files in the CSFII 1994-96,98 file structure cannot be linked to a LifeLine™ 1989-91 CSFII-type file structure.

The complete conversion strategy is presented in the following Appendices to this manual.

Appendix EEE:

This appendix shows the conversion for each commodity listed in the imported database to the LifeLine codes. It lists the following features:

Source Commodity Code: This is the code for the commodity as found in the source data in the ASCII file

Source Crop Group Code: This is the code for the crop group as found in the source data in the ASCII file.

Crop Description: This is the description of the commodity as found in the source data in the ASCII file, and is essentially the same description used in the LifeLine model for the commodity description

#FF: This is the number of Food Forms associated with the commodity as found in the source data in the ASCII file.

FF Code: This is the code for the Food Form of the commodity as found in the source data in the ASCII file.

Source Food Form Description: This is the description of the Food Form of the commodity as found in the source data in the ASCII file. For a complete listing of the Food Forms and the bridge conversion logic, see Appendix FFF.

LifeLine Crop Group Code: This is the LifeLine Crop Group code associated with the source crop group code.

LifeLine Commodity Code: This is the LifeLine Commodity Code associated with the source commodity code

Appendix FFF:
This appendix shows the conversion for all food forms listed in the imported database to the LifeLine food form codes. Note that these conversions are relatively simple because they are defined by the USDA/EPA Translation Files that accompany the USDA CSFII 1994-96, 98 survey.

The conversions provided in these three Bridges are summarized in Figure XXXXXX below.

5.1.3 RECIPES (1989-1991) / TRANSLATION FILES (1994-1996, 1998)

A recipe file (translation file) represents the central link between residues in agricultural commodities and foods as consumed. Consequently, it represents the central element of any analysis of dietary exposure. Originally developed in 1982 by the Office of Pesticide Programs of the US Environmental Protection Agency, these files have in the intervening years been refined and expanded by private consulting organizations. Those refinements reflected the changes in the American food technologies, dietary options and market dynamics. Because of the immense effort involved in maintaining these files, they were high value, proprietary intellectual property. The content of these files were not made available for inspection, even to the licensees of those dietary exposure assessment systems.

LifeLine™ provides full access and inspection of these recipe files (translation files) as part of its dietary exposure analysis system. It is able to do so because the owners of one

of the key sets of recipe files agreed to make them public through this mechanism. The translation files, a product of the Federal government, are obviously publicly available. All future revisions to the recipe files (translation files) in LifeLine™ will also be open and accessible.

A recipe file provides, on a proportional mass basis, the ingredients (specific food forms of Commodities) in a food that was reported eaten in the dietary consumption survey. An example food from the 1989-1991 Continuing Survey of Food Intake by Individuals (CSFII) is shepherd's pie with beef.

2731151 SHEPHERD'S PIE WITH BEEF

An individual record would record the mass of this item consumed by an individual on a particular eating occasion at a specific time on a given survey day.

The corresponding recipe file (Table 5-5) identifies the mass contribution of the food's ingredients and how each was processed.

Table 5-5. Recipe for Shepherd's Pie with Beef (2731151), CSFII 1989-1991

Ingredient	Processing	Mass
Peppers-Sweet (Garden)	Boiled	1.36
Tomatoes-Puree	Canned: Cooked	5.47
Celery	Boiled	2.62
Onions-Dry-Bulb (Cipollini)	Cooked	3.25
Potatoes (White)-Peeled	Boiled	52.37
Wheat-Flour	Canned: Cooked	1.78
Corn Grain-Oil	Cooked	0.32
Cottonseed-Oil	Cooked	0.03
Soybeans-Oil	Cooked	2.08
Sunflower-Oil	Cooked	0.01
Coconut-Oil	Cooked	0.01
Palm Oil	Cooked	0.01
Milk-Nonfat Solids	Cooked	0.22
Milk-Fat Solids	Cooked	0.19
Milk Sugar (Lactose)	Cooked	0.27
Beef (Boneless)-Fat	Canned: Cooked	2.41
Beef (Boneless)-Fat	Cooked	1.24
Beef (Boneless)-Lean (Fat/Free)	Cooked	18.14
Beef (Boneless)-Lean (Fat/Free)	Canned: Cooked	0.17
Pork (Boneless)-Fat	Cooked	0.01
Milk-Based Water	Cooked	2.03
Water-Tap	Cooked	0.55
Water-Commercial Processing	Canned: Cooked	4.93
Water-Commercial Processing	Cooked	0.18

A similar food in the 1994-1996, 1998 CSFII is:

27311510 Shepherd's pie with beef

The corresponding translation file identifies the mass contribution of the food's ingredients. *All are assigned the same CSFFCM status:* ***cooked/not specified/baked****.*

Table 5-6. Recipe for Shepherd's Pie with Beef (27311510), CSFII 1994-1996, 1998

Ingredient	Mass
Potato, tuber, w/o peel	50.12
Beef, meat	16.103
Sugarcane, sugar	0.008
Milk, water	4.324
Milk, nonfat solids	0.49
Milk, fat	2.269
Beef, fat	1.468
Beef, meat byproducts	0.13
Beet, sugar	0.006
Tomato, puree	4.63
Pepper, bell	2.5
Celery	2.73
Onion, dry bulb, dried	0.015
Onion, dry bulb	3.85
Wheat, flour	0.768

Using the residue distributions for each of these ingredients, the software is able to calculate the distribution of residues resulting in the food as consumed.

5.1.4 VARIATION IN TIME AND SPACE

Historically, much of the analysis of dietary exposure, including residues in foods, has been conducted based on annualized national data. While this simplifies the analysis, it overlooks sources of variation in food residues that may be critical to actual exposures. For some commodities, there are major differences in regional source as a function of season, and for some the fraction that is imported varies equally strongly with season.

National, and particularly annual, averaging obscures this variation, and may lead to significant distortions in modeling aggregate exposures.

For example, in winter, table grapes are generally imported. The pesticides used in the growing countries, as well as the use patterns, may vary markedly from U.S. practice.[56] Thus, exposures via foods that contain table grapes may show dramatic seasonal variation.

In order to address this issue, LifeLine™, incorporates the ability to specify residues, use probability factors, and processing factors on either an annual or a seasonal basis. The user, however, may specify either annual or seasonal data, whether entering data manually or in a database file. If both seasonal and annual data are available in a food residue file, the system will preferentially use seasonal data.

Seasonal differences in the processing factors may also be appropriate. During some seasons, a crop may be grown primarily for fresh commodity presentation in the market. During other seasons, the same crop may enter the processed food chain, being prepared as a juice, dried form or in cooked, canned or frozen forms. Some imported foods come into the U.S. market in a specific food form that may be different from the form produced from the domestically produced crop.

LifeLine™ permits the assessor to enter processing factors at the seasonal detail, to correspond to the actual agricultural/market profiles considered for entry of the residue and use data.

[56] Note, however, that pesticides on imported foods must comply with Tolerance limitations, just as domestically produced food. The actual level of that residue, below the tolerance value, or the probability that a residue exists at all, may vary greatly depending on the food source and pesticide use practices at that source. Domestic usage may be equally variable, given the wide range of ecological regions in which some crops are grown.

5.1.5 CALCULATION OF RESIDUES IN FOODS

5.1.5.1 Key Composite Operations

Prior to generating residues for foods (as consumed) the system integrates the three types of information that you have supplied for residues, probability of use and processing factors. Thus, the residue data you entered are combined with data on the probability of use, to generate a complete set of potential residues (including both nonzero and zero values), processing factors are combined into a net factor, that is then applied to scale residues up or down, and generic data (i.e., Crop Group and Commodity data) are applied to all subsets (i.e., Commodities and food forms) where specific data have not been entered.

5.1.5.1.1 Adjusting Residue Distributions for Probability of Use

The assumption made in defining residue distributions for the system is that there are two types of "non-detect" events for pesticide residues that require different strategies. In some cases, the commodity in question was not treated with the particular AI, so that it is reasonable to assume a true zero residue. In other cases, the AI may have been applied, but not detected in subsequent analysis; in these cases, a proxy concentration that reflects the sensitivity of the analytical method is used.

For any data set in which a probability of use (e.g., percent crop treated) has been specified along with residue data, and in which no zero values have been entered for a residue, the system "adds true zeroes" to the residue distribution, to reflect the value assigned to the probability of use. For example, if you have specified five (nonzero) residue values for a Commodity, and have entered a value of 0.10 for probability of use on that Commodity, the system would create a final residue distribution for the Commodity consisting of the five nonzero values and 45 zeroes.

5.1.5.1.2 Modifying Residue Data with Net Processing Factors

While the system allows you to enter as many as six separate processing factors, the important fact for the evaluation of residues in foods is the *net* processing factor that

applies to the particular food form of the Commodity that is used in a food. Accordingly, the system cross-multiplies the processing factors to determine the net factor.

For example, if you specified the processing factors in Table 4-8 for Apples / Canned:Baked, the system would generate a *net* processing factor of 0.675 .

Table 5-7. Processing Factors for Apples / Canned:Baked

Factor	Value
Dehydration	1.25
Washing	0.75
Heating	0.8
Refining	
Storage	0.9
Other	

Note that if a processing factor has not been specified, the model assumes that processing does not affect the residue concentration (default value of 1.0).

5.1.5.1.3 Propagation of Residue Data

Data can be entered at any level of a hierarchy of description: Crop Group, Commodity, or food form. Where data have been entered at the most specific level (food form of a Commodity), they are applied *only* to that specific food form. The most common instance would be processing factors.

Where data have been entered at a more general level (Commodity or Crop Group), they are assumed to apply to *all* specific instances, *unless* specific data have been provided for that specific instance. Thus, for example, if you specify residues (and probability of use) for Apples, these values apply to all of the relevant food forms:

1989-1991	1994-1996, 1998
• Frozen:Boiled	• cooked; fresh; fried
• Frozen:Cooked	• uncooked; fresh; N/A
• Canned:Boiled	• cooked; not specified; baked
• Cooked	• cooked;f fresh; baked
• Canned:Cooked	• uncooked; cured, pickled, smoked, salted; N/A
• Fried	• salad; fresh; N/A
• Baked	
• Raw	
• Canned:Baked	

Unless different values have been specified for one or more of these specific food forms.

Similarly, if you specified residues for Pome Fruits, they would be applied to:

1989-1991	1994-1996, 1998
• Apples-Dried	• Apple, peeled fruit
• Apples-Juice/Cider	• Apple, peeled fruit- babyfood
• Pears	• Apple, dried
• Pears-Dried	• Apple, dried - babyfood
• Apples-Juice-Concentrate	• Apple, juice
• Pears-Nectar	• Apple, juice - babyfood
• Apples	• Apple, sauce
	• Apple, sauce - babyfood
	• Crabapple
	• Loquat
	• Quince
	• Apple, fruit with peel

Again, any value entered for a *specific* instance (e.g., Pears-Nectar, Quince) would take precedence over the generic Pome Fruit value.

Now the Food Residue Translator calculates the appropriate residue distribution (if any) that applies to each potential food form of each Commodity in each Crop Group, considering the residue values you entered, the data on probability of use, and any processing factors that might apply, using the concepts detailed above

5.1.5.2 Calculating Food Residues from Commodity Information

The Food Residue Translator calculates the corresponding distribution of residues in foods as consumed using the following procedure. First the Translator searches all CSFII records for any foods that contains any ingredient form for which a residue or residue distribution has been derived (as described above). Each of these selected foods has one or more ingredients and at least one of these ingredients has a residue or constructed distribution of residues.

Second, a distribution of residues for the selected food is then created, based on the distributions of residues (or single residue values) of the ingredients and the amounts of the ingredients specified in the food recipe. This distribution is created by one of two methods. Where the distribution of residues involves a small number of values the distribution is determined by calculating all of the possible permutations. Where the number of permutations exceeds 10,000, the program uses a Monte Carlo model to determine the distribution of possible residue values in the food.

The resulting distribution of residue levels for each food's residue is then characterized by an empirical cumulative probability density function. This distribution defined by the minimum and maximum residue concentrations, plus the residue concentrations at each of the following percentiles of the distribution of residue levels in the food: 5,10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 87.5, 90,91, 92, 93, 94, 94.5, 95, 95.5, 96, 96.25, 96.5, 96.75, 97, 97.2, 97.4,97.6, 97.8, 98, 98.1, 98.2, 98.3, 98.4, 98.5, 98.6, 98.7, 98.8, 98.9, 99.1,99.15, 99.2, 99.25, 99.3, 99.35, 99.4, 99.45, 99.5, 99.55, 99.6, 99.65,99.7, 99.75, 99.8, 99.82, 99.84, 99.86, 99.88, 99.9, 99.91, 99.92, 99.93, 99.94, 99.95, 99.96, 99.97, 99.98, and 99.99. These values form the Food Residue File Record for that food item.

This pattern of spacing in the cumulative probability density function is designed to emphasize detail at the upper end of the probability distribution. However, the characteristics of the entire residue distribution are preserved since the "draw" from the residue distribution is dictated by a random number generation.

The resulting food distributions are available to view and export for further analysis. This Food Residue File, the product of the Food Residue Translator, can be saved and used in analyses of potential dietary exposures using the *LifeLine Model* Program.

In the *LifeLine Model* Program, the Food Residue File Record is used to select the concentration in a food using the following process.

First, a random number is generated from 0.00 to 100.00 and used to determine the percentile for a food residue on that day for that individual (p_r).

Second, the model then identifies the pair of adjacent percentiles in the empirical cumulative probability density function that are immediately above (p_a) and below (p_b) the randomly selected percentile.

Third, the food residue concentrations associated with p_a and p_b are identified (c_a and c_b).

Fourth, the food residue concentration that corresponds to percentile p_r (c_r) is calculated by a linear interpolation. The formula used is as follows:

$$c_r = c_b + (c_a - c_b) * \left(\frac{(p_r - p_b)}{(p_a - p_b)} \right)$$

This value, c_x, is then used as the residue concentration for that individual for that day. The net effect of this process is that the values of c_r will follow the same distribution as the original distribution of food residues produced in the food residue translator.

5.2 Residential Pesticide Use

Residential pesticide uses fall into two categories. The first category is the residues that occur as the result of the homeowner using a pesticide product or hiring a commercial pesticide applicator. The second category is pesticide residues that the individual encounters that are applied by others. These pesticide residues occur as the result of public use pesticides and use of pesticides on golf courses.

Pesticide residues from the first category[57] are estimated based on information on the frequency of the use of pesticides in residential settings and the level of residues that result from the use. In LifeLine™, the pattern of the use of pesticides is based upon information supplied by the user and data on pest pressures taken from the National Home and Garden Pesticide use Survey. Pesticide residues from the second category are estimated based on user supplied data.

[57] Exposures to the person who applies the pesticide are evaluated in a different manner. See Chapter 5.

Characterizing exposure to an active ingredient presents a number of challenges. A single AI can be used in multiple products to control multiple pests. The products may be used in multiple locations in a residence. A single product may be applied in a number of methods (a granular product may be applied by sprinkling, shaker can, or drop spreader).

In developing the model of residential exposures the LifeLine™ Development Team attempted to create a system that will allow a consistent way of tracking exposures to an AI that result from multiple products, application methods, and application locations. At the same time, the development team sought to allow the user the opportunity to provide all the relevant information on the nature of products, market share, and product specific exposure information. This has been achieved by setting up a hierarchical system where one active ingredient can be used in multiple products. Each product can be used in multiple ways and in multiple locations.

Each product is defined in terms of its probability of use for the control of the specific pest categories used in the NHGPUS. This allows the model to take advantage of the microenvironment-specific data on pest pressure collected in the NHGPUS survey.

LifeLine™ allows the modeling of residential exposures to an AI from pesticides that the homeowner applies and from commercial applicators that treat the residence. Because of the difference in the factors that determine treatment (the home owner responds to a perception of a pest pressure and commercial applicators operate on a fixed schedule) these applications are modeled separately.

5.2.1 MODELING PESTICIDE PRODUCTS AND APPLICATION METHODS

Each AI modeled by the system for residential residues (and exposures) must have at least one End Use Product Equivalent (EUPE), which in turn must have at least one associated Application Method that is applicable to one microenvironment (ME) class.

5.2.1.1 End-Use-Product-Equivalents, Application Methods, and Microenvironment Classes

For many AIs, there are dozens or even hundreds of products that contain the compound. Many of these products are identical and reflect different producers or differences in packaging. When modeling residential exposure with LifeLine™, these products are addressed as a single entry called an End-Use-Product-Equivalent (EUPE). A EUPE is defined as a category of products that are not distinguishable from one another based on any exposure-relevant properties. A EUPE will represent multiple products with distinct product registrations that are substantially the same. For example, if two competing manufacturers had products with identical composition, there is no need to distinguish between the products in an exposure or risk analysis.

The application method defines the specific way that a EUPE is used. As discussed above a single EUPE can have multiple uses. For example, a wetable powder could be used in a variety of pouring or spraying applications. Because the same product could result in different residues and/or applicator exposures, it is important to track the frequency of use of each method.

Residential microenvironments are locations or objects where pests are controlled and where exposures occur. In LifeLine™, microenvironments include different rooms/locations in a residence such as, kitchen, bathroom, garage, lawn and objects such as pools, pets, and clothing[59]. The definition of these microenvironments is based on the survey questionnaire used in the NHGPUS. These microenvironments have been grouped into four classes; indoors, turf, outdoors other, and pets. These Microenvironmental Classes (MCs) were created to better reflect the label instructions for use of EUPEs. For example, products often specify that they are for use indoors or on turf; however, the products rarely state that they are for use in the kitchen but not in the bathroom.

[59] In Version 2.0 of LifeLine™ the only object that is a microenvironment is "pets" future versions may include other objects such as clothing/bedding and foundations.

A given EUPE may have one or several associated application methods, and be used in one or several MCs.

5.2.1.2 Physical Form / Application Method and ME Class

In order to account for the form, application method, and MC, LifeLine™ allows the user to select from one of 55 combinations of these factors. These reflect a combination of:

- Four physical forms (dust/powder, granular, liquid, solid/collar);
- Nineteen application methods (hand, shaker can, bulb or bellows duster, drop spreader, belly grinder, rotary spreader, bomb, aerosol can, hand-pressurized pump sprayer, hose-end sprayer, trigger sprayer, dip, rinse, shampoo, directed-stream spray can, foam spray can, top spot, paintbrush, pour, and collar); and
- Four MC (indoor, turf, outdoors non-turf, and pets).

Because not all application methods apply to all physical forms or MCs, the list of available types does not include all of the permutations of the factors. For example, hose-end sprayers are not used indoors, while top spot treatments and collars apply only to pets.

The system allows the user to assign to each EUPE any and/or all of these application methods. Each application method, however, can only be assigned once for a particular EUPE. If the user has two application methods that differ in their quantitative descriptions (see below) but have the same physical form, application method and ME class applicability, they must be described as separate EUPEs.

5.2.1.3 Pests and Pest-Specific Use Factors

Once the EUPEs/ application methods/locations are defined for an AI, the user provides information on the frequency of use. This is done in two different ways, depending on whether the EUPE is applied by the homeowner or is applied by a commercial applicator.

If the homeowner applies the EUPE, the model requires information on which pests are treated and how likely a specific product/and application method will be used to treat a pest in a ME Class. The system determines the frequency with which a given pest is

treated in a particular microenvironment by examining the NHGPUS record selected for the particular residence being evaluated. For example, the record might indicate that ants were treated three times per year in the kitchen (an indoor ME). The use of these data by the system is described in the next section.

For the system to determine which EUPE/Application Method is employed when a particular pest is treated in a ME, it relies on information supplied by the user on the relative likelihood that specific EUPEs and Application Methods are used to treat that pest in the ME Class of which the particular ME is a member. As noted in the previous section, Version 2.0 of LifeLine™ addresses four distinct ME Classes:

- Indoor
- Turf
- Outdoor Non-Turf
- Pets

Each Application Method for a EUPE is assigned a ***Use Factor*** (UF) ***for a particular pest in a particular*** ME Class. For a particular pest *within* an ME Class, the sum of all UFs must be less than or equal to 1.0. Thus, for example, if there are six EUPE/Application methods that might be used to control ants indoors, and five of them each have a UF of 0.15 (for ants indoors), the UF of the sixth cannot exceed 0.25.

For a different ME Class or different pest, however, the UFs may be completely different, because each UF is specific not only to the pest but also to the ME Class. Thus, in the above example, the sixth EUPE/Application Method might have a UF of 0.5 for ants in Turf, and a UF of 0.7 for cockroaches indoors. The constraint is that the UFs for a particular pest in an ME class must sum to no more than 1.0; UF sums across ME classes or across pests within an ME class may exceed 1.0. This allows the user, for example, to specify one EUPE/Application Method as always being used (UF = 1.0) on ants indoors, while another is always used when the ants are treated on turf, and still another is used on ants in the garden (Outdoor non-turf). Similarly, a EUPE/Application Method may have

a UF of 1.0 for ants indoors, without at all constraining the UFs of other EUPE/Application Methods for cockroaches indoors.

After having specified an Application Method, including physical form of the EUPE and relevant ME Class, the user selects the pest or pests against which this particular application method for the EUPE is employed. Any of the 22 categories of pest for which the system contains data may be selected, and any given application method may be effective against multiple pests. For pets, only fleas or ticks/chiggers may be selected (the other pests on the list do not specifically infest dogs or cats).

For each selected pest, the user must specify a UF that represents the probability that this application method of this EUPE will be used to treat that pest, in the ME Class already selected for the EUPE/Application Method. For example, if a trigger-spray liquid used indoors has 50 percent of the use (roughly equivalent to market share) for the control of ants indoors, the user would enter 0.5 in the data input cell. These data are used in the model to determine the probability that someone with a specific pest problem will use the EUPE/application method to control the pest. The system tracks all of the use factors that have been specified for a particular pest category in an ME class, and ensures that the total is less than or equal to 1.0.

NOTE:
The ***use factor*** (UF) is comparable to ***market share*** **IF** market share is defined in terms of a specific pest and ME Class, **AND** includes all relevant EUPE/Application Methods, whether or not they are included in your exposure and risk analysis. Thus, if a product outside your assessment accounts for 70 percent of the market for the treatment of ants indoors, the sum of UFs for ants treated indoors in your analysis should not exceed 0.30.

The following case illustrates how these data are entered into the system:

There are three EUPEs containing Delta, a hypothetical Active Ingredient that is effective against ants and cockroaches. Product A is a dust/powder that may be used in a bulb duster or drop spreader. Product B is a liquid in an aerosol can for indoor and outdoor use, while Product C is a concentrated liquid that may be diluted for use in a pump sprayer or used in a hose-end sprayer.

Ants and roaches may each be treated in three of the four ME Classes (indoors, turf, and outdoor other). Thus, the user has six pest/MC combinations for which to specify UFs:

- ants, indoors
- ants, turf
- ants, outdoors non-turf
- cockroaches, indoors
- cockroaches, turf
- cockroaches, outdoors non-turf

Not all application methods for the three products are applicable in all ME Classes (drop spreaders are used on turf, hose-end sprayers are used on turf and outdoor other). Thus, the user must specify the following matrix of use factors:

Pest / MC	EUPE / Application Method					
	Product A		Product B	Product C		***Other***
	Bulb Duster	Drop Spreader	Aerosol Can	Pump Sprayer	Hose-End Sprayer	
ants, indoors						
ants, turf						
ants, outdoors non-turf						
cockroaches, indoors						
cockroaches, turf						
cockroaches, outdoors non-turf						

The blocked-out areas in the matrix represent combinations of Application Methods and MC classes that do not occur. For each combination of a pest and MC, the use factors applied to a EUPE/Application method may well differ. Thus, for example, Product A in a bulb duster may be used to treat cockroaches indoors with high probability, but rarely be used to treat ants in any MC. The key constraint imposed by the system is that each ***row*** must total to no more than 1.0.

In this regard, the ***Other*** column in the matrix, while independent of the LifeLine™ system, is critical for the user's development of an accurate exposure assessment. If, for example, 90 percent of the time people treat cockroaches indoors they use a bait that

contains an Active Ingredient other than Delta, the user should make sure that they specify UFs for cockroaches indoors that sum to 0.10. The LifeLine™ software can only check for obvious mathematical errors; it cannot ensure a realistic description of pesticide use.

For commercial application of a EUPE, the model assumes that application events will be determined by the schedule of the applicator rather than the pest pressure, as is the case for resident-applied pesticides. The schedule of the applicator will be driven by the seasonal nature of the pests and the duration of the effectiveness of the products used. NHGPUS tracks whether a residence is served by either indoor or lawn commercial treatment independently of pesticide use by residents.

If the EUPE/application method is commercially applied, the system does not require data on the pests and MEs addressed, or the fraction of treatment for a particular pest. Rather, the user supplies data on the fraction of residences treated with the EUPE / Application Methods by region, for single- and multifamily homes, and the frequency of treatment during warmer and colder seasons for single- and multifamily homes.

Within a season, uses are assumed to occur regularly. For example, if the user specified that the pesticide is used at a frequency of two applications per month, in the South the pesticide would be applied every 15 days around the year. This approach reflects the data available from the NHGPUS survey. (See the following discussion.)

5.2.1.4 Modeling Frequency of Use of Public Use and Golfing Pesticides

Public Use pesticides are those products used by local governments to control mosquitoes, black flies, and other wide spread pests on a neighbor hood basis. The use of these pesticides result in residues on turf and post application exposures to adults and children. The frequency and probability of use of a product near a given residence is a function of region and season. The user enter the fraction of homes treated in a region of the US and if treated the number of treatments per month. The user enters this data into LifeLine using the *AIPD*.

This data is in the form of a fraction of the residential lawns that are treated. The probability that a residence will have a residue from the Public Use will be determined by this fraction. If a home is selected as treated then the home is treated every year. Thus, the model should make the determination of whether a home is treated at each move.

Therefore, when a person moves to a new house the model should determine if there is a lawn (based on the AHS data.) If there is a lawn, then the model determines if the house is in a treated neighborhood. If the house is in a treated neighborhood then the house will always be treated.

When an application occurs a new residue is added to the existing residue (if any) on the turf. This residue is then treated like any other residue, see Section 5.2.3.

Exposure to residues from golf occurs when the modeled individual plays at a local course (located in the same census region as his or her residence). The probability that the course will have a residue is determined using the following process.

At age 12, the LifeLine™ module should begin the daily update of the residue level on the golf course ME. On the first day of an individual's 13th year of life (they have just turned 12), the model determines if the golf course that the modeled individual plays on a user of the product? (Note the model should determine if the person is a golfer. If the person is not a golfer then skip this source of exposure.) This is determined using the Use factor entered in the *AIPD*. If the product is not used, then ignore this source of exposure until the individual's next move. Once a person moves then a new golf course is assumed to be used and LifeLine™ will determine if the pesticide is used on that course.

5.2.2 MODELING PEST PRESSURE

Each time an individual is assigned to a new residence (whether at birth or because of a subsequent move), the pest pressure for that residence is determined based upon data from the NHGPUS.

5.2.2.1 NHGPUS

NHGPUS characterized the use of pesticides in U.S. residences. The survey was designed to be representative of the U.S. as a whole. This survey provides a record of the yearly use of pesticides in the residence of the person interviewed. Each record presents the pest pressure (number of times treatment was required for a specific pest) by residential location and the application method (crack/crevice, broadcast, bomb, etc.). LifeLine™ assumes that while the occurrence of a pest varies from day to day, the pest pressure is stable over time and is a function of the home. Thus, the record of the number of times a pest required control is likely to be relatively constant over time (little or no year-to-year variation) for a residence. Accordingly, the NHGPUS record for a home is selected for a new residence and is assumed to hold for the total duration of time an individual lives in the home.

Each record of pesticide use in a residence specifies the type of home (detached or multiple family), setting (urban or rural), census region, and presence of a yard. The software bins the NHGPUS records according to these characteristics. After a first or new home has been selected (See Chapter 3), a NHGPUS record is selected randomly from the bin that matches these characteristics in that home. In this fashion, the characteristics of the home assigned in the mobility process and from the AHS data are consistent with the set of homes from which the NHGPUS record was selected.

Once the NHGPUS record is selected, then additional characteristics are established for the residence. These include:

- Presence of one of the following: fruit trees, nut trees, or grape arbor;
- Presence of a vegetable garden;
- The use of a commercially applied pesticide indoors; and
- The use of a commercially applied pesticide out of doors.

5.2.2.2 Development of Daily Probabilities of Pesticide Use for Homeowner Applied Pesticides

Once the NHGPUS record is selected, the pesticide use history of the home is used to predict the daily probability of using a homeowner-applied pesticide for the entire time

the person lives in the house. The history specifies the number of applications that occurred in the survey year in a residential location to control a specific pest. The dates on which these treatments occurred are not specified. As a result, temporal patterns of use cannot be directly determined from the record.

In Version 2.0 of LifeLine™, it is assumed that the probability of using the pesticide in a single-family residence is the same on all days when the outdoor temperatures are relatively warm. In the Midwest and the Northeast, this is assumed to occur for only half of the year, in the West for three-quarters of the year, and year-round for the South. This assumption is not made for multifamily dwellings. The larger building sizes for multifamily residences are thought to make the pest pressures for many pests to be less dependent on seasonal effects.

The daily probability of use is defined as the probability of using a pesticide at a particular location to control a specific pest. The model uses a binomial function to determine the use or lack of use for each day. The specific application method used to control the pest is *not* tracked from the NHGPUS record, but is assigned using the approach described below. While the NHGPUS data are presumably representative of enduring features of pest pressure, it is unlikely that they are reflective of the current market share of different pest control product types.

Under this approach, the number of pesticide applications for a particular pest and microenvironment may vary slightly from year to year in any residence. This occurs because the number of days that will be modeled as having pesticide use will not be exactly the same as the "Reported number of applications" in NHGPUS. Instead, the number of uses will follow binomial distribution with a mean equal to the "Reported number of applications" in the NHGPUS record. This procedure, however, is consistent with the assumption that the "Reported number of applications" is the result of a constant pest pressure that in a given year resulted in the observed number.

5.2.2.2.1 Minimum Repeat Time (MR)

A limitation with the approach described above is that the basic model assumes that the use of a pesticide is independent of the use on prior days. This assumption is not reasonable, use

of a product should greatly reduce the probability of use on subsequent days. It is also inconsistent with label requirements on many products. Accordingly, the model adds an additional step in the determination of the daily frequency that sets the probability of use equal to zero for a specified number of days after each use. This number of days is called the minimum repeat time (MR).

The value for MR is based in information on the label. For example, if a pesticide label says "Repeat as needed," the minimum time is zero. If the label says, "Repeat every seven days" then the minimum repeat time is seven days. During this period, the model will assume that the pesticide is not used (probability of use is zero on the six days following each use). In order to determine time since prior use, the model sets a counter that is reset to 0 after each use. On each subsequent day, the counter is increased by one. When the counter value is less than the MR for the specified EUPE and application method, the probability of use is zero.

This calculation must be determined for each application method that is used in the microenvironment to control the pests indicated by the user. See below for a description of the assignment of particular products and application methods within a residence.

5.2.2.2.2 Daily Probability of Pesticide Use in a Micro Environment

Based on the above assumptions, the daily probability of using a pesticide to control a specific pest at a specific location (microenvironment) is given by taking the number of reported uses and dividing by the number days in seasons likely to have pest problems minus the average number of days that will fall into the exclusion periods immediately after applications (MR). The equation used for single-family residences is as follows:

$$DP = RA / ((FY * 365) – (RA * MR))$$

Where:

DP is the probability of using a pesticide application on a given day of the portion of the year when pests are active.

RA is the reported number of applications is taken from NHGPUS (for a specific pest, at a specific location) over one year.

MR is the minimum repeat time specified for a particular EUPE and application method by the model user.

FY is the fraction of the year that pesticides are most likely applied.

The determination of FY is described below.

5.2.2.2.3 Seasonal Limitation (FY)

For single-family homes, the daily probability for all regions is set at zero for seasons with cold temperatures, and at one for seasons with warm temperatures. Specifically the Northeast and the Midwest are assumed to have no pest pressure from mid-fall to mid-spring. The West is assumed to have no pest pressure during winter, and the South is assumed to have pest pressure year-round. As a result, the value of FY will be 1.0 for the South, 0.5 for the Midwest and the Northeast, and 0.75 for the West.

5.2.2.2.4 Pesticide Use Locations

The NHGPUS identifies use locations that correspond to multiple microenvironments (Living Room/Bedroom/Nursery/Den, Other Indoor Areas). In LifeLine™, the use of pesticides is assumed to be simultaneous in all of these locations. Thus, if the pesticide is modeled as being used in a bedroom, it is also assumed to be used in the same way on the same day in the individual's den and living room.

NHGPUS also reports the frequency of the use of a product in a location (bathroom, bedrooms, etc.) that may correspond to multiple rooms in many homes. Homes may have multiple bathrooms, multiple bedrooms, and multiple "other indoor areas". Thus, use of a pesticide in "a bedroom" may not result in an exposure if the modeled person

uses another bedroom. In Version 2.0, the assumption is made that all bathrooms and all bedrooms are treated simultaneously.

These two assumptions will result in overestimation of exposure to an AI. The impact of alternative assumptions should be evaluated in future versions of the model.

5.2.2.3 Assigning Pest Pressure to Particular Products

As noted above in the description of pesticide products (EUPEs) and application methods, each EUPE/Application Method combination has a particular probability of being used for a specific pest in a class of microenvironments (Indoors, Turf, Outdoor Non-Turf, and Pet). The total probability assigned to all products for a combination of a particular pest and ME class cannot exceed 1.0. (It may well be less than 1.0. For example, a competing product not addressed in a particular risk analysis may have significant market share.)

The model addresses the potential use of multiple EUPES or application methods by the following plausible simplification: Each year, the entire probability of treating a particular pest in a particular ME class is assigned to a single EUPE/Application Method.[60] For that entire year, each treatment is assumed to use the identical product. This approach provides a better approximation to reality for many pest control situations than to assume that the choice of product is random for each event. This may be an overly conservative[61] assumption that forces the modeled individual to use one product for an entire year when some products may be changed more frequently.

5.2.2.4 Commercially Applied EUPEs / Application Methods

Unlike the case for homeowner-applied pesticides, the NHGPUS provides only minimal data on the use of commercial pesticides. The record only indicates whether the home or yard is commercially treated. Because of the limited data available for commercially

[60] If the modeled EUPEs have a total probability less than 1.0, in some years no treatment will occur (i.e., treatment is assumed to have occurred with a competing agent that is outside the defined risk group).

applied pesticides, LifeLine™ requires additional inputs from the user on frequency of use (by season and type of residence). The limitation of data also poses a problem for determining which of the indoor microenvironments are treated commercially. Version 2.0 makes the conservative assumption that pesticides are applied to all rooms in a household during commercial treatment.

5.2.2.5 Pest Strips

The one exception to the above approach is pest strips. Pest strips are linked to pest pressures (fabric insects or other pests); however, the strips are assumed to be used prophylacticly. That is as one strip is exhausted it is replaced with a fresh strip. The result is that once a house is determined to use pest strips they are assumed to be continually used until the person moves to a new home.

5.2.2.6 Pet Ownership

NHGPUS contains data on whether pets, pet living areas, or pet bedding were treated, without distinguishing between these. Neither does NHGPUS identify the nature of the pet(s) in a residence. LifeLine™ makes two simplifying assumptions in modeling exposure via pets:

- All treatments occur to the pet, and all exposure is mediated by the pet; and
- All pets are either cats or dogs.

When modeling EUPEs/application methods for pets, the user must specify whether the product is meant to treat a cat or dog. The model assumes that all dogs weigh 30 lbs and that all cats weight 10 lbs. These choices are used to control the quantitative exposure to the pesticide following application. For a given mass applied to an animal, treatment of a dog results in a lower concentration of AI per square meter than does treatment of a cat. In contrast, a dog offers a larger surface area for exposure than does a cat.

---

[61] In general, the greater consistency that is imposed upon an exposure assessment, the higher will be the upper percentile distributions of exposure for a modeled population.

In Version 2.0 of LifeLine™, however, when the record indicates the presence of a pet, it is randomly assigned to be a dog or cat.

The further assumption is that pets are linked to residences (because the data indicate that pet treatments are specific to the residence). When an individual is assigned to a home with pet treatment, the pet is classified as a dog or cat. The pet remains constant for the duration of residence. When the individual moves, the presence of a pet is independently determined for the new residence.

This pattern holds for the duration of time that the person lives in the home (i.e., it is a function of the home). A drawback of this approach is that pet ownership is linked to the home not the person. As a result, the person becomes or ceases to be a dog owner when he or she moves into a new home. This occurs because the data on pet ownership reported in the NHGPUS are specific to homes.

5.2.2.7 Determining the Frequency of Use for Public Use Pesticides

As discussed in section 5.2.1.4, of LifeLine™ uses user-supplied data to determine if a residence receives residue for public use pesticides. Once the house is determined to be treated, then the daily probability of the application of a new residue is determined. The daily probability will be zero for the cold weather portions of the year. For the warm periods of the year, the applications will be determined by the frequency of application data given in the *AIPD*. That data will be used in the following fashion.

First, when a person moves to a new home and the house is determined to receive residues from a public health use of the product, the model will determine if that day is a "warm day" i.e., occurs during the warm portion of the year. If the day occurs in the warm portion of the year, then an application is assumed to occur on that day. If it is not during the warm portion of the year, then the model waits until the first day of the warm portion of the year and then assumes that an application is performed.

Second, the model assumes that the applications are made on a regular cycle. The period between the applications is determined based on the user-supplied data on frequency

(applications per month). Assuming that a month has 30 days, the period between the applications is given by:

Period = 30/Frequency

At the end of the period, a second application is made. This is repeated until the end of the warm season. At the beginning of the next warm season the process begins again with a new initial application. This continues until the person moves. If the warm season does not end (in the South) it goes on until the person moves.

5.2.2.8 Determining the Frequency of Use for Pesticides Applied to Golf Courses

LifeLine™ uses user-supplied data to determine if a person plays golf and if his or her golf course uses the pesticide. Once this is determined the probability that a pesticide is applied on a given day is determined using the following process. If the product is used, then determine if this is a warm or cool day using the day of the year and the region. Based on the warm/cool status of the day the mode determines the probability of using the product using the equations given below.

If Season/region is cool then:

DP=0

If Season/Region is warm then:

DP = RA / ((FY * 365) – (RA * MR))

Where:

DP is the probability of using a pesticide application on a given day of the portion of the year when pests are active.

RA is the annual number of applications for the golf course taken form the .rkg.

MR is the minimum repeat time specified for a product in the .rkg.

FY is the fraction of the year that pesticides are applied.

The value of FY will be 1.0 for the South, 0.5 for the Midwest and the Northeast, and 0.75 for the West. The region the person lives in will be used to select the value of FY for his or her golf course.

Finally, if the product is used, then set the minimum repeat clock to determine when the product is next used.

5.2.3 OTHER PESTICIDES

The pesticides that are addressed in Version 2.0 are primarily defined by the NHGPUS survey. This survey excluded pesticides used in agricultural production, plant growth regulators, pool chemicals, and anti-fouling paints. As a result of this data limitation, this version of LifeLine™ does not include information on these pesticides. Nor does NHGPUS account for pesticides applied in the common areas of multi-unit residences. Thus, pesticides used by the operators of apartment complexes for outdoor areas (including children's play areas) are not included in the survey results.

5.2.4 POST-APPLICATION RESIDUES

The calculation of residues immediately following an application, and on subsequent days, reflects properties not only of the AI, but also more importantly of the specific EUPE(s) and application method(s) that are employed. Additional parameters are collected on the AI that relate to exposure via residues in tapwater, the calculation of dose from exposure, and the estimation or risk (See Chapters 6 and 7).

5.2.4.1 Applications and Residues

The same basic approach used to determine the residues of a pesticide Active Ingredient (AI) in each of the media (surfaces, air, soil, pet, or grass) in each of the indoor and outdoor microenvironments (MEs).

The system begins the modeling of the residues for each day by determining whether one or more user-specified End-use-Product-Equivalents (EUPEs) were used in the ME. If no EUPEs are used, then the residues on the relevant medium in a ME is estimated based on the level in the ME on the prior day using the following equation:

$R_{current} = R_{prior} * (100\text{-}PD)/100$

Where,

$R_{current}$ is the residue on the surface of an ME on the current day.

R_{prior} is the residue on the surface of an ME on the prior day.

PD is the user-supplied value for the daily percent decline in either the dislodgeable residue (PD_{Dis}), the total residue (PD_{Tot}), or both, as appropriate to the specific ME and exposure scenario (See Chapter 7).

If a pesticide is used then:

$R_{current} = NR + R_{prior} * (100\text{-}PD)/100$

Where,

NR is the incremental concentration of residue in the media associated with the new application.

5.2.4.2 Dislodgeability and Decline

In addition to the mass of AI that is applied in any given EUPE / Application Method, two factors drive the residues on surfaces[62] that are available for post-application exposure. The first of these is dislogability. The incorporation of a dislodgeability factor in LifeLine™ reflects the assumption (well-supported by empirical evidence) that a fraction of the mass of any AI that reaches a surface will be bound to the surface in some fashion, and will not thereafter be available for exposures. Only the unbound (dislodgeable) fraction is relevant to many exposure scenarios.

The use of a single, fixed fraction for dislodgeability for a given EUPE / Application Method probably represents a major oversimplification of the relevant kinetics between the AI (and other ingredients in the EUPE) and whatever surface(s) are being modeled.

[62] Airborne residues are addressed separately, as described below.

The second factor is the decline rate, which is specified for each EUPE and Application Method. For MCs (e.g., turf) where both dislodgeable and total residues are important for potential exposure, the user is asked to separately specify a decline rate for each.
Low decline rates may be associated with significantly elevated exposures in situations where a EUPE is frequently reapplied. Leaving the decline rate at zero will result in a rapid increase in residues over successive applications, and will wildly overestimate post-application exposures.

5.2.4.3 Limitations on the General Equation

Two factors modify the general form of these equations for daily residue. The first limits the calculation of daily decays in residues to those concentrations that are likely to be meaningful for exposure analysis. The other controls the probability that the user will encounter post-application residues as calculated on the day of treatment.

5.2.4.3.1 De Minimis Level (DML)

As discussed above, the user can supply a De Minimis Level (DML) for each AI that is modeled. LifeLine™ uses the user-supplied DML to limit the calculation of trivial residues that result from residue decline in the days following application. For example, if the surface residue was found to be less than the DML the concentration on a given day would be set to zero, and would remain at zero until a pesticide was used in that ME. If the user sets the *de minimis* value to zero, this process will not occur and the equation to evaluate daily declines in residues will be run no matter how low the concentration. The software uses the presence of the zero concentrations to skip over media and MEs that are of little concern, improving the speed of the model.

5.2.4.3.2 Before and After Use

This step determines the probability that the time the individual spends in a ME occurs prior or following the application of the pesticide (P_{BA}). The value of P_{BA} is set by the user between 0 and 1. A value of 0.0 means that all pesticide uses will occur after to the individual spends time in the ME on a given day and that the individual will not be exposed to post-application residues on the day of the application. A value of 1.0 will result in the model assuming that the EUPE was applied immediately before the

individual enters the ME and the individual's exposures will be affected. A value of 0.5 results in the applications happening prior to the use for 50% of the time and 50% after.

The model compares a randomly selected number (between 0 and 1.0) to P_{BA}. If the value is greater than P_{BA}, then the pesticide application is assumed to occur after to the time spent in the ME and the residues for that day's exposures are calculated without the contribution of that day's use. Thus, the residue equals the prior day's residue levels corrected for decline is used. However, the total residue level for the ME (post-application) is still calculated for use in the estimation of the following day's residues. If the value is greater than P_{BA}, then the residue does include the contribution from the current day's use.

The value of P_{BA} only affects the estimation of post application exposures.

5.2.4.4 Medium-Specific Equations

The following sections present the specific equations that apply to media in indoor and outdoor MEs.

5.2.4.4.1 Indoor Surfaces

In the case of indoor surface concentrations of the AI, LifeLine™ focuses on the dislodgeable fraction of the surface residue in each ME. The following equations apply:

On the day of application:

$$DR_{indoor\ surf} = (NRS * FD) + (DR_{prior\ surf} * (100\text{-}PD_{dis})/100)$$

Where,

$DR_{indoor\ surf}$ is the dislodgeable residue concentration on the surface of the ME on the current day (mg/m^2).

NRS is the new residue on a surface and is defined as the amount of AI per area of surface applied on the current day (mg/m^2).

FD is the fraction of the AI that is dislodgeable (unitless).

$DR_{prior\ surf}$ is the dislodgeable residue concentration on the surface of the ME on the prior day (mg/m^2).

PD_{Dis} is the percent decline in the dislodgeable residue that occurs over one day (unitless).

On the following day (assuming no additional application):

$$DR_{indoor\ surf} = DR_{prior\ surf} * (100 - PD_{dis})/100$$

The value of the FD will differ for hard and soft surfaces. However, the same equations will be used.

5.2.4.4.2 Pest Strips and Indoor Air

Pest strips are solids that are designed to continually emit pesticides to the air. The air concentration will be determined by the rate of emission from a strip, the house size and the residential air exchange rates. However, data on the emission rates were not available as a result an alternative approach was used that is based on reported monitoring data associated with pest strips.

The first day air concentration will be provided by the user in the AIPD. This air concentration will occur ***in all indoor microenvironments*** not just in the ME where

NHGPUS indicates the pest were controlled. If a pest strip is used in a home, then the value of $AC_{average}$ for <u>all</u> indoor MEs is given by the following equation:

$$AC_{averagek} = AC_{PS} * e^{-k*24*n}$$

Where,

AC_{PS} is the value of the initial air concentration entered in the AIPD for the pest strip or pulled from the distribution.

k is the first order constant of the decline in the airborne concentration in units of $hours^{-1}$ provided by the user in the AIPD, and

n is the number of days since the pest strip was placed in the closet. (On the day that the pet strip is placed in the closet n=0.)

5.2.4.4.3 <u>Products Other than Pest Strips and Indoor Air</u>

Indoor air concentrations for products other than pest strips is based only on the releases to the air that occur during application, as a result the decline in residues in air is controlled not by the product- and application-method-specific decline rate, but rather by the air exchange rate for the modeled residence[63]. Because air exchange rates are large for most residences, and because the model assumes that air exchange involves the infiltration of air that is free of residues, the decline in air concentrations following a pesticide application is substantial. This rapid decline has a significant impact on the average air concentration that an individual would encounter if they spent a long period of time in an ME following an application.

In the software, we have assumed that the air concentration will follow a first order decline. Under such an assumption the average air concentration over a period of time D is given by:

[63] Version 2.0 of LifeLine™ does not consider air concentrations that occur from the volatilization of pesticides after application. Future versions of the model may consider this additional source of airborne residues.

$$AC_{average} = AC_{initial} \frac{(1 - e^{-ER*\mathrm{ETjk}})}{ER * \mathrm{ETjk}}$$

Where,

$AC_{average}$ is the average concentration of the AI in air (mg/m^3).

$AC_{initial}$ is the initial concentration of the AI in air (mg/m^3).

ER is the air exchange rate (hr^{-1}).

ET_{jk} is the duration of time spent performing the j^{th} activity in the k^{th} microenvironment (hr)

The value of $AC_{average}$ will be used in the calculation of dose from inhalation (Chapter 6). The value of $AC_{initial}$ is calculated in the following way. On the day of application:

$$AC_{initial} = AC + (AC_{prior\ initial} * (1-e^{(-24*AE)}))$$

Where,

AC is the estimate of the air concentration that will occur indoors as result of the use of product at the end of the use or at the end of the period of exclusion specified on the label (mg/m^3).

$AC_{prior\ initial}$ is the initial concentration calculated for the prior day.

On the following day (assuming no additional application):

$$AC_{initial} = AC_{prior\ initial} * (1-e^{(-24*AE)})$$

5.2.4.4.4 Turf and Soil

Because of the potential oral ingestion of grass, the equations for turf are modified from those presented above. The equation for dislodgeable residues on turf (DR_{turf}) (used for dermal and grass to hand to mouth exposures) will be the same as the indoor surface levels:

On the day of application:

$$DR_{turf} = (NRS * FD) + (DR_{prior\ turf} * (100\text{-}PD_{dis\ turf})/100)$$

Where,

DR_{Turf} is the dislodgeable residue concentration on the surface of the turf on the current day (mg/m^2).

NRS is the new residue on a surface and is defined as the amount of AI per area of surface applied on the current day (mg/m^2).

FD is the fraction of the AI that is dislodgeable (unitless).

$DR_{prior\ truf}$ is the dislodgeable residue concentration on the surface of the turf on the prior day (mg/m^2).

$PD_{dis\ turf}$ is the percent decline in the dislodgeable residue that occurs over one day (unitless).

On the following day (assuming no additional application):

$$DR_{turf} = DR_{prior\ turf} * (100 - PD_{dis\ turf})/100$$

The level on the turf (R_{Turf}) (used for oral ingestion of grass) will be given by these alternative equations:On the day of application:

$$R_{turf} = NTR + (R_{prior\ turf} * (100\text{-}PD_{turf})/100)$$

Where,

R_{Turf} is the total residue concentration on the surface of the turf on the current day (mg/m^2).

NTR is the new residue on a surface and is defined as the amount of AI per area of lawn applied on the current day (mg/m^2).

$R_{prior\ truf}$ is the total residue concentration on the surface of the turf on the prior day (mg/m^2).

PD_{turf} is the percent decline in the total residue that occurs over one day (unitless).

On the following day (assuming no additional application):

$$R_{turf} = R_{prior\ turf} * (100 - PD_{turf})/100$$

Neither dermal nor hand-to-mouth exposure pathways are modeled for soil. There is; however, a direct ingestion pathway. This pathway will use the total residues in soil (R_{soil}). The total residue in soil is determined by making the conservative assumption that all of the applied AI will be retained in the top centimeter of soil. In a one square meter area the volume of the soil in the top one centimeter will be 0.01 m^3 or 10,000 cm^3. Assuming a soil density of 1.5 g/cm^3 this corresponds to a soil mass of 15,000 g of soil. The value for R_{soil} is given by the following equations:

On the day of application,

$$R_{soil} = (NSA / SM) + (R_{prior\ soil} * (100\text{-}PD_{soil})/100)$$

Where,

R_{soil} is the total residue concentration in the soil on the current day (mg/kg).

NSA is the new soil application of AI and is defined as the amount of AI per area of surface applied on the current day (mg/m^2).

SM is the mass of the soil in the top one centimeter of a one square meter area of lawn (15 kg).

$R_{prior\ soil}$ is the total residue concentration in the soil on the prior day (mg/g).

PD_{soil} is the percent decline in the total residue in soil that occurs over one day (unitless).

On the following day,

$$R_{soil}\ (mg/g) = R_{Soil}\ (mg/m^2) * (100 - PD_{soil})/100$$

5.2.4.4.5 Golf Course Residues

Exposure to residues only occurs when an individual plays golf. These exposures occur as a result of dermal contact with the turf on fairways, tees, greens, or roughs. However, the determination of residues requires that the model track the residues on each day of a golfer's life independent of whether they play or not. Residues do not need to be tracked if the person is less than age 12 or is a non-golfer.

The dislodgeable residue on the golf course turf is determined in the same way as the dislodgeable turf residues. On the day of application:

$$DR_{turf} = (NRS * FD) + (DR_{prior\ turf} * (100\text{-}PD_{dis\ turf})/100)$$

Where,

DR_{Turf} is the dislodgeable residue concentration on the surface of the turf on the current day (mg/m^2).

NRS is the new residue on a surface and is defined as the amount of AI per area of surface applied on the current day (mg/m^2).

FD is the fraction of the AI that is dislodgeable (unitless).

$DR_{prior\ turf}$ is the dislodgeable residue concentration on the surface of the turf on the prior day (mg/m^2).

$PD_{dis\ turf}$ is the percent decline in the dislodgeable residue that occurs over one day (unitless).

On the following day (assuming no additional application):

$$DR_{turf} = DR_{prior\ turf} * (100 - PD_{dis\ turf})/100$$

5.2.4.4.6 Gardens

The dislodgeable residue on the foliage of a garden is determined in the same way as the dislodgeable turf residues. On the day of application:

$$DR_{gard} = (NRS * FD) + (DR_{prior\ gard} * (100\text{-}PD_{dis\ gard})/100)$$

Where,

DR_{gard} is the dislodgeable residue concentration on the garden foliage on the current day (mg/m^2).

NRS is the new residue on a surface and is defined as the amount of AI per area of surface applied on the current day (mg/m^2).

FD is the fraction of the AI that is dislodgeable (unitless).

$DR_{prior\ gard}$ is the dislodgeable residue concentration on the surface of the garden foliage on the prior day (mg/m^2).

$PD_{dis\ gard}$ is the percent decline in the dislodgeable residue that occurs over one day (unitless).

On the following day (assuming no additional application):

$$DR_{gard} = DR_{prior\ gard} * (100 - PD_{dis\ gard})/100$$

5.2.4.4.7 Pets

Exposure to AI residues from pets occurs from the dislodgeable residues. Accordingly, the equations for pets are substantially the same as those for turf and indoor surfaces.

On the day of application:

$$DR_{pet} = AA * FD_{pet} + DR_{prior\ pet} *(100 - PD_{Dis})/100]$$

Where,

DR_{pet} is the dislodgeable residue concentration in the fur of a pet on the current day (mg/m^2).

AA is the amount of AI per unit of surface area applied to the pet on the current day (mg/m^2).

FD is the fraction of the applied AI that is dislodgeable from the pet's fur.

PD_{dis} is the percent decline in the dislodgeable residue that occurs over one day.

On the following day:

$$DR_{pet}\ (mg/m^2) = DR_{pet}\ (mg/m^2) * (100 - PD_{Dis})/100$$

5.2.4.5 Incremental Residues from Specific Application Methods

This section presents the equations and assumptions used to estimate the incremental residues associated with the application of a pesticide product (an EUPE with a particular

application method) based on data entered by the user. The general approach used in these equations is to use simple models of the residue levels that are consistent with the EPA Residential Exposure SOPs but that take advantage all of the data provided by the prior characterization of the individual, her residence, and her behaviors. In particular, the models take advantage of the data on air exchange rates, room types, and sizes.

5.2.4.5.1 Estimation of Residue Levels from Use of a Bomb

A bomb is an aerosol container that is intended to be "set off" and to fully discharge its contents without the handler or any other individual being present in the room. The amount of AI released from a can is determined from the label. The user enters the amount of AI in a unit and the size of the room each unit treats. The model assumes that the minimum number of cans necessary to treat each room(s) will be used. Bombs are always assumed to be used indoors.

Surface Residues: The model estimates the amount of AI released to the floor (NRS) based on the amount of AI in a can and the number of cans used. The concentration on the floor and other contact surfaces is given by the mass of AI released divided by the room's floor area.

$$NRS = UA * NU / RS$$

Where,

UA is the mass of AI in a bomb,

NU is the number of units set off in a room, and

RS is the size of the floor area of the room where the pesticide is applied.

This estimate of NRS is conservative because some of the AI will be lost to the outdoors and to ceilings and walls.

The value of NU is defined by rounding up to the next whole number the following equation:

$$NU = RS / A$$

Where,

A is the area (floor area) treated by one unit.

Air Concentration: The amount of AI in the air is a user input that should reflect the amount of AI that is present in the ME at either the end of the application, or end of the exclusion period (if specified on the label), in the form of either a vapor or an aerosol sufficiently small that it behaves like a vapor. The units for air concentration are mg/m^3.

5.2.4.5.2 Estimation of Residue Levels from a Broadcast Application Indoors

Indoor broadcast applications are assumed to be applied to floors and other contact surfaces directly, with an application mass that is directly proportional to the area to be treated.

Surface Concentration: This application method is assumed to be used only indoors. The user provides an estimate of the amount of AI that is applied per square meter in the form of a value for the amount of the product applied and the fraction of the product that is AI. In the absence of any prior residues:

NRS = FA * AP * FT

Where,

FA is the fraction of the AI in the product (unitless).

AP is the amount of product applies per floor area (m^2).

Air Concentration: The amount of AI in the air is a user input that should reflect the amount of AI that is present in the air of the ME at either the end of the application, or end of the exclusion period (if specified on the label), in the form of a vapor or in an aerosol sufficiently small that it behaves like a vapor. The units for this are mg/m^3.

5.2.4.5.3 Estimation of Residue Levels from a Spot Treatment Indoors

Spot treatment is essentially similar to broadcast treatment indoors, with the exception that only a (generally small) fraction of the total area of the room is treated.

Surface Concentration: The user provides an estimate of the amount of AI that is applied per square meter. The user also provides an estimate of the fraction of the floor on the ME that is treated (FT). This value will be used as the value for $R_{indoor\ surf}$. In the absence of any prior residues:

NRS = FA * AP * FT

Where,

FA is the fraction of the AI in the product (unitless).

AP is the amount of product applies per floor area (m^2).

FT is the fraction of the floor area of a room treated (unitless).

The model assumes that the area(s) that have been spot treated have the same probability of being contacted by the individual as the untreated portion of the room.

Air Concentration: The amount of AI in the air is a user input that should reflect the amount of AI that is present in the ME at either the end of the application, or end of the exclusion period (if specified on the label), in the form of a vapor or in an aerosol sufficiently small that it behaves like a vapor. The units for this are mg/m^3.

5.2.4.5.4 Estimation of Residue Levels from Crack and Crevice

Crack and crevice treatments are assumed to be applied to the perimeter of a room, with residues applying to all surfaces. The total mass applied will be a function of the rate of application of the AI and the room's perimeter. In order to derive an estimate of residue on the floor the applied mass will be assumed to be spread equally over the entire area of the room.[64] This is a conservative assumption that will result in overestimate of the available residues. As a result, the user may wish to reduce the estimate of the amount applied to better estimate the potential for post application exposures.

[64] Alternately, one can assume that the residues are considerably higher towards the perimeter of the room than in the center, but that the potentially exposed person is equally likely to spend time in the room (regardless of activity) in the areas of high residue as in those with lower residues.

Surface Concentration: The concentration on the surface from the use of a crack and crevice is estimated based on the following equation.

NRS = FA * AP * RP / RS

Where,

FA is the fraction of the AI in the product (unitless).

AP is the amount of product applied per meter of crack (mg/m).

RP is the perimeter of the room (m).

RS is area of the floor of the room (m^2).

Air Concentration: The amount of AI in the air is a user input that should reflect the amount of AI that is present in the ME at either the end of the application, or end of the exclusion period (if specified on the label), in the form of a vapor or in an aerosol sufficiently small that it behaves like a vapor. The units for this are mg/m^3.

5.2.4.5.5 Estimating Residue Levels from Use of Indoor Fogger

An indoor fogger is designed to treat the entire volume of a room, rather than the surfaces within it. In this, it resembles a bomb, with the exception that the applicator is present when the application is made.

Surface Concentration: The concentration on the surface from the use of an indoor fogger is estimated based on the following equation.

NRS = FA * PA * RH

Where,

FA is the fraction of the AI in the product (unitless).

AP is the amount of product applied per volume of air (mg/m^3).

RH is the room height in meters for a room with a ceiling of 8 ft (2.44 m).

Air Concentration: The amount of AI in the air is a user input that should reflect the amount of AI that is present in the ME at either the end of the application, or end of the exclusion period (if specified on the label), in the form of a vapor or in an aerosol sufficiently small that it behaves like a vapor. The units for this are mg/m^3.

5.2.4.5.6 Estimating Residue Levels on Grass and in Soil from a Broadcast Application on Turf

When broadcast application of a EUPE occurs in an outdoor ME, it is assumed that there is sufficient dispersion to rapidly reduce the air concentration to trivial levels. Therefore, no estimates of air concentration are made.

5.2.4.5.6.1 Soil Levels

The amount of AI applied to an area of soil (NSA) is given by:

$$NSA = FA * PA$$

Where,

FA is the fraction of the AI in the product (unitless).

AP is the amount of product applied per area of lawn (mg/m^2).

NSA is amount of AI that is applies as a result of an application (mg/m^2).

5.2.4.5.6.2 Level on Turf

The level of residues on turf (NTR) is defined by the model user's inputs on the EUPE. The data is used in the following equation.

$$NTR = FA * PA$$

Where,

FA is the fraction of the AI in the product (unitless).

AP is the amount of product applied per area of lawn (mg/m^2).

NTR is the new soil residue that occurs as a result of an application (mg/m^2).

5.2.4.5.7 Estimating Residue Levels on Pets

No air residues are assumed to result post application from treatment of pets. The residue on the pet is defined as the applied amount (AA) divided by the surface area of the pet.

$$AA = FA * PA / PS$$

Where,

FA is the fraction of the AI in the product (unitless).

AP is the amount of product applied per pet (mg/animal).

PS is the surface area of the pet in m^2.

5.3 Tapwater Concentrations

Tapwater residues are selected from four seasonal distributions, which are matched to each residence based on Census region, urban or rural setting, and the type of water supply (public or private system, individual well, or other) specified in the AHS record for the house (see Chapter 3).

First, a house is assigned randomly to a percentile. Then, the corresponding seasonal value is selected from the seasonal distributions for that region, urbanization, and water supply type. For example, if a residence has tapwater from a public or private water supply, and is located in an urban area in the Northeast, and it has been randomly assigned to the 12^{th} percentile, then the values corresponding to the twelfth percentile for each season's distributions assigned to the home's water supply.

This procedure assures that if there is no seasonal variation in water levels, then the concentration of the AI in tapwater will be constant for a residence. In addition, if there are seasonal variations, a residence that has an upper or lower percentile residue in one season will have a correspondingly high or low residue in all other seasons. Finally, the same four residue levels will be repeated each year that the person resides in the home.

Inhalation exposures due to the presence of residues in tapwater involve modeling the volatilization of those residues in a shower (this implementation of the model does not address general contamination of household air by substances that volatilize from tapwater, which may be associated with significantly greater exposures). The equations used for the volatilization of residues in a shower are described in Chapter 6.

5.3.1 DEVELOPING DATA ON TAPWATER CONCENTRATIONS

The following guidance is offered for the development of distributions

Public and private water systems can obtain water from surface waters or wells or a mixture of both. In general, smaller rural water supplies will tend to rely on wells and larger urban supplies on surface water supplies. However, the fraction of water derived from surface and wells may be extremely complicated and vary by the location in a water district and season. The model assumes that the user will define the distribution of concentrations for this category of water supplies to reflect the distribution of contaminants in the taps of homes in urban and rural areas.

Residential wells are most likely to be affected by local uses of pesticides containing the AI under analysis. As a result, the number of affected wells will often be limited to residences near the areas where the products containing the AI are used. This suggests that information on the fraction of the population living in counties where an AI is applied may be useful for developing a distribution.

Based upon screening models, monitoring data or other factors, the user may have already determined that an AI 1) is only a concern for ground water supplies, 2) is only a concern for surface water supplies, 3) is a concern for both, or 4) is not a concern for either supply. Based on these finding the user may chose to "turn off" the tapwater related exposure routes, limit the tapwater exposures to private wells, or model exposure from multiple water supplies.

Information on where and when a pesticide is used can also be used to determine the distribution of concentrations of the AI in residence's water supplies. If the use of a pesticide is limited to certain regions or counties then the tapwater concentrations can be set at zero for the remaining regions.

5.3.2 ENTERING DATA USING CUMULATIVE DISTRIBUTIONS

The Tapwater Concentrations program enters the data as an empirically defined cumulative distribution (the fraction of a population that is exposed to a concentration equal to or less than a specific value). This distribution is defined in terms of a minimum concentration a maximum concentration and sets of concentrations and percentiles. (The minimum and maximum concentrations can be thought of as the concentrations that correspond to the 0^{th} and 100^{th} percentiles of the cumulative distribution). The model then interpolates between these values to describe the total distribution. This program produces a file with the ending *.twc*. The **.twc* files are imported in the *LifeLine Model* when tapwater related exposures are calculated.

Entering data using the format can be confusing. The following are instructions for entering commonly encountered distributions.

5.3.2.1 Entering a fixed value:

The number of percentiles should be set to zero. The minimum and maximum value should be set to the same value.

5.3.2.2 Entering a uniform distribution:

The number of percentiles should be set to zero. The minimum and maximum value should be set to the minimum and maximum value of the uniform distribution.

5.3.2.3 Entering an empirical distribution:

The number of percentiles will be defined by the user based on the minimum number necessary to characterize the distribution. The minimum and maximum value should be set to the minimum and maximum concentrations anticipated to occur.

5.3.2.4 Entering a parametric distribution:

using Excel or other statistical software package values for an appropriate number of cumulative percentiles of the distribution should be determined and entered into the table. Many distributions are unbounded. The model requires that the user set minimum and maximum values. These values can be based upon theoretical considerations such as solubility or other considerations. The current version required that this distribution be entered by hand. Future versions will allow the direct importation of data files.

5.3.2.5 Entering a mixed function distribution:

A mixed function distribution is one where the distribution of concentrations is the result of mixtures of separate populations each with its own distribution. This will be a common distribution for tapwater concentration where one population of individuals will typically have zero residues (no pesticide use in the region where they live) and another where there is a distribution of residues. Let the fraction of a region's population that has zero residues be 0.80 and let the cumulative distribution for the remaining fraction have the following form, (Min =0.05, Max = 1.5, P_{25} = 0.1, P_{50} = 0.5, P_{95} = 1.0). The distribution would be entered as follows:

The minimum concentration would be zero and the maximum 1.5;

The percentiles for the distribution would be 4;

The first percentile that would be entered would be, P_{80} = 0;

This would result in the model selecting a value of zero 80% of the time. The remaining three percentiles will have the same concentrations but the values of the percentiles will be changed using the following equation.

$$P_r = P_o (1 - P_1) + P_1$$

Where,

P_r is the revised percentile.

P_o is the percentile from the original population.

P_1 is the fraction of the population that had a zero exposure.

Therefore in this example the remaining three percentiles would be (P_{85} = 0.1, P_{90} = 0.5, P_{99} = 1.0). This will have the model select for the population with non zero tapwater residues 20% of the time.

CHAPTER 6. MODELING EXPOSURE AND DOSE

As a model of aggregate exposure, LifeLine™ estimates exposures from multiple sources and by multiple routes. LifeLine™ also evaluates the absorption of AIs by each route and provides estimates of the absorbed dose. The sum of the absorbed doses from the different routes provides the basis for the estimates of aggregate doses and the estimates of aggregate risks. Separate estimates of dose are performed for each microenvironment and activity that and individual performs in a given day.

Estimating exposure and dose requires different approaches for different sources. This section presents a description of the equations and approaches used to estimate exposure to AIs and the resulting dose from dietary, tapwater, and residential sources of AI exposure.

6.1 Contact Rate and Intensity

In general, exposures by all routes and from all sources are quantified as the combination of a contact rate (e.g., how many slices of pie did a person eat in a week, how often did a child touch a surface with a pesticide residue in an hour?) and an intensity (what was the residue in the pie or on the surface, how much residue was transferred from the surface to the hand?). Chapter 4 discusses many of the variables that influence contact rate and intensity, while Chapter 5 addresses the prediction of residue concentrations that are available for contact.

Dietary exposure has the simplest equation of all the sources of exposure:

Exposure = Mass of Food Item Ingested x Residue in Food Item

In dietary exposure assessments, the complexity of the analysis occurs in the selection of the appropriate dietary record and in the calculation the probable distribution of residues in the foods in that record. These procedures are addressed in Chapters 4 and 5.

The basic equations for post-application exposures in residential settings are more complex. These equations are taken from the December 1998 Draft Residential SOPs developed by EPA's Office of Pesticide Programs. The equations have been modified to include terms not specified in the SOP, in order to allow the user to investigate higher tiered assessments and to support the use of chemical, demographic, behavioral, and residential information supplied by the user or generated by the software.

Equations for exposure from tapwater are based on current EPA guidance. The models of inhalation exposure from volatilization of residues also have a long history of use in exposure assessments.

This chapter addresses the equations used to combine the descriptions of behaviors discussed in Chapter 4 with the descriptions of residues discussed in Chapter 5, to yield actual estimates of exposure and/or dose. In LifeLine™, all exposures and doses are calculated on a mass per day basis. Once the total doses from all sources are summed by route (and by all routes) they are divided by the individual's body weight to give the final estimates of exposure/dose on a milligram of Active Ingredient (ai) per kilogram of body weight per day basis. Handler exposures are calculated using an alternative approach and are addressed separately in a subsequent section of this chapter.

6.1.1 GENERAL FORMS OF EXPOSURE EQUATIONS - RESIDENTIAL (POST APPLICATION) AND TAPWATER

While the specific equations used to address exposure in any given scenario (e.g., dermal contact with turf residues following a turf treatment) are unique, the fundamental equations that are used have the same general form for each route of exposure. Specific variables control for unique features of a behavior (e.g., hand-to-mouth) and unique characteristics of a microenvironment - ME (e.g., dislodgeable residues on hard surfaces).

6.1.1.1 Inhalation Exposure (post application)

Version 2.0 of LifeLine™ only evaluates inhalation exposures for pesticide residues for indoor microenvironments. The equation used in the current draft of the SOPs has been used as the basis for the evaluation of the dose from indoor inhalation of an AI:

$$\text{Inhalation Dose}_{jk} = AC_{\text{average } k} * IR_j * ET_{jk} * LC$$

Where,

Inhalation Dose_{jk} is the dose (mg/day) from inhalation of airborne AI while performing the j^{th} activity in the k^{th} microenvironment.

$AC_{\text{average } k}$ is the average airborne concentration of the AI in the k^{th} microenvironment (mg/m^3).

IR_j is the inhalation rate for the j^{th} activity (m^3/hour).

ET_{jk} is the duration of time spent performing the jth activity in the kth microenvironment (hr).

LC is the fraction of the mass of AI that is cleared from the lung (not exhaled).

The inhalation rate for an activity reflects both the user-modifiable activity class assigned to the activity (resting, sedentary, light, moderate, or heavy) and the physiology (age, height, and weight) of the modeled individual, see Chapter 2. The derivation of airborne concentrations in indoor microenvironments from application of pesticides by various methods is described in Chapter 5.

Where a value of 1.0 is used for LC the model will produce an estimate of inhalation exposure (the mass that enters the lung), where the actual lung clearance for an AI is used, the model will estimate the absorbed dose by inhalation.

6.1.1.2 Dermal Exposure

Different equations are used to calculate dermal exposure and dose for residues on surfaces and for residues in water. In the case of contact with residues on surfaces (or dermal exposure during application, discussed below), the historical approach has been to

first determine dermal exposure and to separately consider absorption and dose. LifeLine™ takes this approach, providing the user the option of determining the amount that reaches the skin (exposure) or the amount absorbed (dose).

A very different approach has been used for dermal exposure to dilute aqueous-phase AIs. There, exposure assessment methods have all addressed dose directly, taking consideration of dermal absorption in the structure of the equations used to assess exposure. The reason for this is that a model of a loading of contaminant on skin does not work well with the constant flush/refresh regimen of the shower, or the essentially infinite theoretical source term for dermal contact during swimming. These exposure situations are really driven by partitioning between the skin and water, rather than a simple determination of the amount of AI that reaches the skin.

6.1.1.2.1 Direct Dermal Contact

Potential dose rates from dermal contact with a residue resulting from a specific activity (as defined by the NHAPS survey) in one microenvironment are calculated as follows:

$$\text{Dermal Dose}_{jk} = DR_{\text{surface } k} * TC_j' * SA * ET_{jk} * CF_j * DAF$$

Where,

Dermal Dose_{jk} is the dose rate (mg/day).

$DR_{\text{surface } k}$ is the dislodgeable residue level on the surface (mg/cm^2) of the k^{th} microenvironment.

TC_j' is the age- and activity-specific transfer coefficient normalized to the individual's surface area (hr^{-1}).

SA is the surface area of the individual (cm^2).

ET_{jk} is the duration of the behavior in the k^{th} microenvironment (hr).

CF_j is the age- and activity-specific clothing factor, and

DAF is the compound or product specific dermal absorption factor.

The particular dislodgeable residue level that is relevant varies as a function of the microenvironment, while the transfer coefficient is a user-modifiable property of the relevant activity. The DAF can be defined for an AI or it can be made product specific.

When DAF is set at 1.0, the equation produces the dermal exposure (mass of AI reaching the skin) when DAF is set to the AI's actual dermal absorption , the equation will estimate the absorbed dermal dose.

6.1.1.2.2 Water-Borne Exposure

LifeLine™ uses equations for dermal exposure to aqueous-phase contaminants developed by EPA (EPA, 1992) .The fundamental form of the equation is:

$$Dose = DA_e * SA$$

Where,

DA_e is the dose absorbed per unit area of exposed skin per event (mg/cm^2).

SA is the surface area of the entire body of the exposed individual (m^2).

The specific equations used to address the absorbed dose in a shower are described below. In Version 2.0, only whole-body doses from residues in water are addressed.

6.1.2 TRANSFER BETWEEN MEDIA IN A MICROENVIRONMENT

The general equations presented above assume that appropriate residue data for the relevant medium are available. In the case of inhalation exposure from residues in tapwater, these equations must be supplemented by the prediction of concentrations in air that result from the domestic use of water with a known residue concentration. Subsequent enhancements of the model may require additional information in intermedia transfers or residues.

6.1.2.1 Volatilization

In order to assess inhalation exposures to residues in tapwater, it is necessary to predict the concentrations in air that will be produced by introducing a residue of a particular AI into the domestic water supply. This entails a choice of models. For example, there are models of volatilization into general household air from all domestic water uses, and there are models of more specific situations, notably of volatilization into shower stall or

bathroom air during a shower event. In keeping with the room- and activity-specific modeling used throughout LifeLine™, the volatilization model (and the associated exposure model) specifically address showering.

The simple volatilization model used in the system is derived from the work of Andelman, and is well fit to empirical studies of volatilization (Giardino et al., 1990). Most of these studies, however, have been concerned with highly volatile compounds (high Henry's Law coefficients). In Andelman's original formulation, a simple screen was used to establish whether volatilization in the shower occurred at all, for AIs with a unit-less Henry's coefficient below 10^{-4}, it was assumed that no volatilization occurred. Other AIs were assumed to have significant volatilization.

This formulation is not as well fit to data on semi volatile compounds such as AIs. While it is clear that Henry's law coefficient is not very predictive of the differences in volatilization that are seen for volatile compounds (unitless Henry's coefficient > 10^{-1}), volatilization is less than complete for compounds with lower Henry's Law coefficients, but still above Andelman's original proposed cutoff of 10^{-4}. More complex models (such as those of McKone and Little) note a proportionality of mass transfer from water to air with the Henry's law coefficient, which reaches an asymptote at values of the coefficient between 10^{-2} and 10^{-1} (unitless).

The approach taken here is to assume that the volatilization fraction is essentially proportional to the mass transfer coefficient, below the asymptote predicted for that coefficient. That is, concentrations in the shower are not limited by steady state partitioning between bathroom air and water. Rather, conditions in the bathroom are such that concentrations in bathroom air achieved during bathing events continue to increase, proportional to the concentration of the AI in the water, the water flow rate, and the AI-specific volatilization factor that reflects the mass transfer coefficient.

6.1.2.1.1 Predicting Air Concentrations from Tapwater Concentrations

The concentration in the bathroom air is calculated using Andelman's model (Andelman, 1990) for the shower:

C_{air} = $((C_{water} * fv * Fl_{water} * t_s) / V_{bath}) * 1.5$

Where,

C_{air} is the Average air concentration during and after shower (mg/m^3).[65]

C_{water} is the tapwater concentration in residence at that season (mg/ml)

fv is the fraction of the contaminant volatilized (derived below).

Fl_{water} is the Flow rate of water from the shower (l/hr) [Default at 700].[66]

t_s is the time spent in the shower (hr), equals one-half of time spent in the activity (*Personal hygiene - bathing etc.* Code 40).

V_{bath} is the volume of the selected bathroom for the individual (m^3).[67]

This model *over*estimates air concentrations, due to lack of air infiltration into the bathroom, while it *under*estimates air concentrations due to the assumption of perfect mixing in bathroom air. Future versions of the model may account for air exchange rates[68].

6.1.2.1.2 Rules for Assigning Volatilization Fraction:

The following rules are used to predict the degree of volatilization of a tapwater residue into bathroom air, based upon user-supplied data for the AI:

- For Henry's Law Coefficient (Unitless[69]) > 10^{-1}, volatilization is asymptotic (90% assumed volatilized in a shower).

---

65 This reflects the assumptions, 1) the activities described as *personal hygiene, bathing, etc.* are equally divided between showering and post-showering and 2) the concentration of the contaminant in air reflects a constant loading proportional to water flow, starting at zero and reaching a final level when the water is turned off

66 Subsequent versions may include an option for varying flow rates as a function of showerhead design. This is at the high end of the range of reported data.

67 As discussed in Chapter II, the system provides bathrooms that are either 400 or 800 cubic feet (11.3 or 22.7 m^3). For those residences where there is more than one bathroom, individuals over 18 get a large bathroom, while those under 18 get a small bathroom

68 For example, assuming the shower runs for half of the time the person does the activity, the equation for the average air concentration over time t would be as follows: Cair = 0.5* Cwater *fv * Flwater * (ts – (1-e-kats)/ka + (1-ekats)2/ka) Vbath * ka *ts where, ts is ½ of the bathing time, and ka is the air exchange rate for the residence.

69 The unitless form of Henry's is derived from the form that specifies units by the universal gas law equation. This reduces to the simple proportion H = H' * 41 where H' is Henry's Law Coefficient in atm-m^3 / Mol.

- For Henry's Law Coefficient (Unitless)< 10^{-3}, volatilization is proportional to Henry's Law Coefficient. And is given by the following equation:

  $$vf = 90 * H,$$

  Where,

  vf is the fraction volatilized (unitless).

  H is the Henry's Law Coefficient (unitless).

Thus, for:

$H = 10^{-3}$, vf = 0.09 (9 percent)

$H = 10^{-4}$, vf = 0.009

$H = 10^{-5}$, vf = 0.0009

etc.

- For $10^{-3} < H < 10^{-1}$, a linear approximation of the asymptote is used. This is crude (one would expect a supralinear function for the asymptote seen in the more complex shower models), but well within the range of observational variability:

  $$vf = 0.09 + 8.18 (H-10^{-3})$$

This yields the predictions given in Table 6-1.

Table 6-1. Predicted Volatilized Fraction (vf)	
H	**vf**
0.001	0.09
0.003	0.11
0.010	0.16
0.030	0.33
0.100	0.90

6.2 Scenario-Specific Equations

The general equations presented above have a series of specific forms that address particular patterns of behavior in different types of microenvironment. Variants of the basic equations exist for:

- Different modes of non-dietary oral exposure,
- Exposure via pets, and
- Exposure to residues in tapwater.

6.2.1 NON-DIETARY ORAL EXPOSURE

Three different forms of non-dietary oral exposure to residues have been addressed in EPA's SOPs and Version 2.0 of LifeLine™: hand-to-mouth behavior, ingestion of turf (grass), and ingestion of soil. The calculation of CR_j, and the selection of R_k, differs for each source of oral exposure.

6.2.1.1 Hand-to-Mouth

Hand to mouth exposure occurs when an individual contacts a contaminated source with their hand and then place their hand or a portion of their hand into their mouth. The behavior is believed to occur most frequently in children and is influenced by the activity the individual is performing. Activity specific values are required for many of the inputs to the exposure/dose model. This behavior occurs in indoor MEs and outside on turf. Potential dose rates from non-dietary ingestion of a residue of an AI on an ME surface associated with a specific activity (as defined by the NHAPS survey) being performed in a microenvironment are calculated as follows:

$$NDOral\ Dose_{jk} = DR_{indoor\ surf\ k} * CR_j * ET_{jk} * GIA$$

Where,

$NDOral\ Dose_{jk}$ is non-dietary oral dose of an AI (mg/day) for the j^{th} behavior in the k^{th} residential microenvironment.

$DR_{indoor\ surf\ k}$ is the dislodgeable residue level on the relevant surface in the k^{th} residential microenvironment (mg/m^2).

CR_j is the contact rate associated with the j^{th} behavior. Different behaviors will use different parameters to calculate contact rate.

ET_{jk} is the duration of the j^{th} behavior in the k^{th} residential microenvironment (hr).

GIA is the fraction absorbed from the gastrointestinal tract.

When GIA is set at 1.0 the equation produces the oral exposure (mass of AI entering the mouth), when the fraction of absorption is entered as GIA, the equation produces an estimate of the absorbed dose.

The contact rate for non-dietary oral exposure is given by the following equation:

$$CR_j = MF * FH * R * EC * OMF$$

Where,

MF is the number of time per hour an individual places some portion of their hand in their mouth (mouthing frequency) (events/hr).

FH is the average fraction of the hand that is placed in the mouth (unitless).

R is the replenishment factor (unitless).

EC is the extraction coefficient (unitless).

OMF is an oral modifying factor (unitless).

The replenishment factor can be thought of as the ratio of the average concentration of an AI on the hand at the beginning of each mouthing event to the dislodgeable concentration on the surfaces of the ME. Where a hand is repeatedly placed in the mouth between contacts with a contaminated surface, the ratio will be small. Where the hand comes in contact with a contaminated surface between each mouthing events the value will approach 1.

The extraction coefficient is the fraction of the AI on the portion of the hand that enters the mouth that is extracted by saliva. Recent guidance from EPA (EPA, 1999), suggests that a default value of 0.5 could be used for this factor.

The oral modifying factor can be a single number or a cumulative distribution. The factor can be used to account for the impact of variation in one or more of the equation's parameters.

6.2.1.2 Turf Ingestion

Turf ingestion is the oral dose that results from incidental ingestion of grass. The exposure only occurs when the individual is in the yard. Potential dose rates from ingestion of soil are calculated as follows:

$$NDOralT_{jk} = R_{turf} * IgR_T * ET_{jk} * GIA$$

Where,

$NDOralT_{jk}$ is the dose from the consumption of turf = grass (mg/day).

R_{turf} is the total residue on grass (mg/cm^2), see Chapter 5.

$IgRT_j$ is the age- and activity-specific ingestion rate of grass (cm^2/hr), see Chapter 4.

ET_{jk} is the duration of the exposure (h/day).

GIA is the gastrointestinal absorption factor

6.2.1.3 Soil Ingestion

Soil ingestion is the oral dose that results from incidental ingestion of soil. The exposure only occurs when the individual is in the yard. Potential dose rates from ingestion of soil are calculated as follows:

$$NDOralS_{jk} = R_{Soil} * IgRS_j * ET_{jk} * GIA / 1{,}000{,}000$$

Where,

$NDOralS_{jk}$ is the dose from the consumption of soil (mg/day),

R_{soil} is the total soil residue (mg/kg), see Chapter 5.

$IgRS_j$ is the age- and activity-specific ingestion rate of soil (mg/h), see Chapter 4.

ET_{jk} is the duration of the exposure (h/day).

GIA is the gastrointestinal absorption factor.

The factor of 1,000,000 converts for units of soil ingestion and residue (milligrams to kilograms)

6.2.2 DERMAL AND NON-DIETARY ORAL EXPOSURE FOR PETS

While the general form of the exposure equations is the same for pets as it is for both hand-to-mouth and dermal exposures to residues on surfaces, the evaluation of these exposures differs in two ways. First, the duration of exposure is not based upon the NHAPS activity record, but rather on the duration specified by the user. The reason for this is that contact with pets is not tracked by the NHAPS. Interaction with pets can occur at any location in a home and tends to be a series of short-term events. Second, the transfer coefficient for pets is defined by both the area of the person that comes in contact with the pet and the size of the pet.

The activity parameters that control contact rates and are entered in the *Activity Descriptions* program, just as they are for activities that are sampled from the NHAPS records. The duration of contact, however, is determined by the user. These data are input in the in the *Analysis Preferences - Residential* tab in the main *LifeLine Model*. The occurrence of the activity, however, is solely reflective of the presence of a pet. If there is a pet in the residence, the "playing with pets" activity will occur every day. The location of the activity is not associated with any specific location in the home since pets

can interact with an individual at any location in a residence. This duration is not accounted for in terms of a specific time of day nor are the durations of other activities adjusted for this period of time. The reason for this is that playing with a pet is assumed to occur as a series of short-term events that occur while the person is performing other activities (watching TV, studying, resting, etc.).

The default value is two hours per day and applies to all ages. This value is based on EPA policy and is thought to be a high-end estimate. Future versions of LifeLine™ may link the duration to the period of time a person spends at home or may allow the user to enter a distribution of times spent.

6.2.2.1 Dermal Exposure

The Transfer Coefficient for dermal exposures, unlike the case for other MEs, reflects both the size of the exposed individual and the size of the pet.[70] The Transfer Coefficient for pets, TC_j' is defined as the smaller of two quantities:

- One- fourth of the surface area of the cat or dog per hour[71] (See Chapter 5), or
- The fraction of the individual's SA specified in the *Activity Descriptions* file.
  Default values of the latter are:
  - 0 for children under age 1 (infants are assumed to have minimal contact with pets)
  - 0.25 for children ages 1-12
  - 0.05 for individuals over age 12

This approach is based on the concept of a child "hugging or sleeping" with a pet and an older individual petting the animal.

[70] For other MEs, the available surface of the individual will always be small relative to the available surface in the ME. This is not the case for pets.

[71] This value is defined as follows: Surface Area (cm^2) =12.3 * (Pet weight (g))$^{0.65}$

6.2.2.2 Hand-To-Mouth Exposure

Estimating hand-to-mouth oral exposure also involves a small variation from the standard equation. The equation used for pets is:

$$NDOralHM_{pet} = DR_{pet} * SA_H * EC * FH_j * RR_j * FQ_j * ET_{jk} * * OMF * GIA$$

Where,

$NDOralHM_{pet}$ is the oral dose from pet- hand-to-mouth exposures (mg/day) for pets.

DR_{pet} is the dislodgeable residue level on the pet's fur (mg/cm^2), see Chapter 5.

SA_H is the surface area of the individual's hand (cm^2), see Chapter 2.

EC is the compound-specific extraction coefficient (removal of the AI from the hand by saliva), see Chapter 5.

FH_j is the average fraction of the hand that enters the individual's mouth for playing with a pet, see Chapter 4.

RR_j is the replenishment rate.

FQ_{pet} is the frequency of hand-to-mouth activity for playing with a pet (events/hr), see Chapter 4.

ET_{pet} is the user-specified duration of pet contact (hr).

OMF is the oral modifying factor (unitless) that can be entered as a single number or as a cumulative distribution.

GIA is the fraction absorbed from the gastrointestinal tract.

6.2.3 EXPOSURE TO WATER-BORNE RESIDUES

As noted above, two special considerations affect exposures to residues in tapwater. Assessing inhalation exposures requires the prediction of air concentrations from water concentrations. Dermal exposures inherently involve less than complete absorption, based on the partitioning of the AI between the aqueous phase and a lipid phase (i.e., skin).

6.2.3.1 Oral Exposure

The selected concentration in the water for the residence and season (See Chapter 5) is multiplied by the consumption of water (See Chapter III) to yield an oral tapwater dose:

$$Dose_{OT} = C_{water} * V_T$$

Where,

$Dose_{OT}$ is the oral dose from tapwater (mg/event).

C_{water} is the concentration of the AI in water (mg/l).

V_T is the volume of water ingested (l), see Chapter 5.

For this model, we assume pure water, such that one liter weighs one kilogram.

6.2.3.2 Inhalation

Inhalation dose is calculated from the modeled air concentration {see above), the age- and activity-specific inhalation rate for the individual, and the duration of the event:

$$Dose_{IT} = C_{air} * I_{ST} * t$$

Where,

$Dose_{IT}$ is the inhalation dose from tapwater from showering (mg/event).

C_{air} is the concentration of the AI in bathroom air (mg/l).

I_{ST} is the inhalation rate during the shower and dry-off period (l/h), (i.e., age- and activity-specific inhalation rate).

t is the time spent in the activity (hr).

The value of t is taken from the NHAPS activity *Personal hygiene - bathing etc. Code 40.*

6.2.3.3 Dermal Exposure/Dose

LifeLine™ addresses dermal exposure from residues in water by modeling exposure during a shower. Slightly different forms of the same equations could be used to address a bath (e.g., fixed volume of water, different time parameters, less-than complete mixing

or skin contact with water), but these would be expected to have relatively small influence on predicted exposure/dose.

As noted above, the calculation of dermal exposure from aqueous residues inherently addresses the mass absorbed from the water, rather than addressing the loading of skin with a mass. Absorbed dermal dose in the shower is as follows:[72]

$$ADDose_{shower} = DA_e * SA$$

Where,

$ADDose_{shower}$ is the absorbed dermal dose from showering.

DA_e is the dose absorbed per unit area of exposed skin per event (mg/cm^2).

SA is the total surface area of the exposed individual (cm^2).

The derivation of Dose absorbed per unit area per event has two forms:

$$DA_e = 2 * K_p * C_{water} * ((6 * \tau * t_e) / \pi)^{1/2}$$

for $t_e < t^*$[73]

$$DA_e = K_p * C_{water} * ((t_e / (1 + B)) + (2 * \tau ((1 + 3B) / (1 + B))))$$

for $t_e > t^*$

Where:

K_p is the dermal permeability of the AI from water (cm/h).

C_{water} is the concentration of the AI in water (mg/cm^3).[74]

---

72 These equations apply only to organic compounds. There are available equations for inorganic compounds, but their applicability to modern pesticides is limited.

73 Based on the distribution of t* found by EPA, it will hardly ever be the case that $t_e > t^*$ so the first equation will likely hold true

τ is the lag time for diffusion of the AI through the skin, reflecting diffusion path and diffusivity coefficient of the AI (h).

t_e is the duration of the event (h).

t* is the time required to reach steady-state [h] (reflecting τ and B)

B is the relative contribution of permeability coefficients for the stratum corneum and the viable epidermis, proportional to lipophilicity

In the software we have assumed that t_e equals 1/2 of the time spent in *Personal hygiene - bathing etc. Code 40.*

Dermal permeability reflects the basic physical/chemical properties of the AI specified by the user:

$$Log(K_p) = -2.72 + (0.71 * \log (K_{ow})) - (0.0061 * MW)$$

Where,

K_{ow} is the octanol-water partition coefficient.

MW is the Molecular weight.

The lag time for diffusion (τ) reflects both the properties of the AI specified by the user and the properties of the skin (model defaults developed by EPA are used):

$$\tau = l_{sc}^2 / 6 D_{sc}$$

Where,

l_{sc} is the path length through the stratum corneum (10 microns).

D_{sc} is the diffusivity of the substance within the stratum corneum (cm^2/h).

Diffusivity in the stratum corneum reflects both path length and molecular weight:

$$Log (D_{sc} / l_{sc}) = - 2.72-(0.0061 * MW)$$

---

[74] The equations from EPA's document use these units. This is 0.001 times the concentration in mg/l.

In turn, the time required to reach steady state reflects the lag time and the lipophilicity of the AI:

$t^* = 2.4 * \tau$

for $B <= 0.1$

$t^* = (8.4 + 6 * \log(B)) * \tau$

for $0.1 < B < 1.17$

$t^* = 6 *(b-(b^2-c^2)^{½}) * \tau$

for $B >= 1.17$

Where,

$b = ((2 * (1 + B)^2)/ \pi) – c$

$c = (1 + 3B) / 3$

Lipophilicity in dermal absorption is assumed to be well-predicted by the octanol-water partition coefficient entered by the user:

$B = K_{ow} / 10{,}000$

6.3 Handler (Applicator) Exposures

The estimation of applicator (handler) exposures in the Version 2.0 of LifeLine™ follows historic practice in EPA's Pesticide Program of estimating the application dose in terms of some fraction of the amount of AI residue that is applied. Two alternative approaches are available, both of which relate handler exposure directly to the application rate for each application method. The first approach is to obtain unit dermal and inhalation exposures from the Pesticide Handlers Exposure Database (PHED) as recommended by the draft Residential SOPs. For users who do not have access to PHED, or for application methods that are not addressed by PHED, it is also possible to specify inhalation and dermal exposures as a percentage of the amount of AI that is applied. The software does not evaluate handler exposures where a pesticide is applied professionally

or where the application method is a bomb (the applicator is not in the room during the application).

Version 2.0 of LifeLine™ requires that the user enter data on handler exposure in the form of a single value for each EUPE and application method. Future versions will allow the user to enter a distribution of values.

As discussed below, the amount of pesticide applied is determined based on the room sizes of the residence and the user supplied application rates. As a result, applicator exposures will vary across individuals even though a single unit exposure or percentage of applied dose is used.

6.3.1 PHED BASED APPROACH

The PHED database allows one to select a set of studies containing measured exposure data from studies of mixers/loaders and applicators[75]. The user can select from data files that address one or more of these populations from which to extract data, and then can select studies on the basis of information describing the study, AI, application site, procedures used to mix, load, or apply, weather conditions, or available data. After selecting an appropriate set of studies, and adjusting for the presence/absence of protective clothing, the user can request a report the exposures that represents the mass of dermal exposure, inhalation exposure, or both. Exposures can be reported either as raw data or normalized on the various bases of time. LifeLine™ uses data that have been normalized based on total mass of AI applied.

Exposures are reported separately for inhalation (in units of nanograms of AI inhaled per pound applied) and dermal exposure (in units of micrograms of AI on the skin per pound applied). These can be directly entered into the description of a EUPE and Application Method for a ME (see Chapter 5). The values of the unit exposure factors can be

[75] Flaggers, who signal cropdusters as to where to apply a pesticide, are also covered by PHED, but are not a population relevant to Version 2.0 of LifeLine™.

described in terms of point estimates or as cumulative distribution of inter application variation in the unit exposures.

The doses are simply the PHED exposure fractions of the applied mass of AI. For inhalation:

$$\text{Inhalation Dose}_{\text{Applicator}} = UE_I * AM * LC / 453{,}593{,}490{,}000$$

Where,

$\text{Inhalation Dose}_{\text{Applicator}}$ is the inhalation dose for that application (mg).

UE_I is the unit inhalation exposure (ng/lb AI).

AM is the amount of AI used in a specific application of a pesticide (mg).

LC is the lung clearance factor (fraction of the AI that is not exhaled).

The constant adjusts from pounds applied to milligrams applied, and from nanograms of exposure to milligrams of exposure.

The dermal exposure equations also include a clothing-modifying factor. This factor, which can be entered as a distribution, or a single factor can be used to characterize the effect of variation in dermal exposure. The factor can be adjusted for hot and cool portions of the growing season.

For dermal exposure:

$$\text{Dermal Dose}_{\text{Applicator}} = UE_D * AM * *MF *DAF / 453{,}593{,}490$$

Where,

$\text{Dermal Dose}_{\text{Applicator}}$ is the dermal dose for that application (mg).

UE_I is the unit exposure (µg/lb AI).

AM is the amount of AI used in a specific application of a pesticide.

MF is the season-specific modifying factor.

LC is the lung clearance factor (fraction of the AI that is not exhaled).

The constant adjusts from pounds applied to milligrams applied, and from micrograms of exposure to milligrams of exposure.

6.3.2 FRACTION OF THE APPLIED AMOUNT

In the absence of a PHED value, the user may directly specify both inhalation and dermal exposures as a percentage of the applied amount of AI. The preset dermal exposure fraction is 10%, with the exception of pet collars where the fraction is 1%. The preset value for inhalation is always 0. These numbers represent EPA policy used in recent reregistration decisions. The user has the ability to change these values.

$$\text{Dermal Dose}_{\text{Applicator}} = \text{FAM} * \text{AM} * \text{DAF}$$

Where,

$\text{Dermal Dose}_{\text{Applicator}}$ is the dermal dose for that application (mg).

FAM is the fraction of applied mass that ends up on the handler (unitless).

AM is the amount of AI used in a specific application of a pesticide (mg).

DAF is the dermal absorption factor.

6.3.3 AMOUNT OF AI APPLIED

Applicator exposures, as indicated above, are calculated using a user-specified term (either a unit exposure value or a simple fraction of the amount of AI applied) and the amount of AI used during each application. The amount of AI applied reflects two factors: the application rate (e.g., mg/m, mg/m^2, mg/m^3) and the perimeter, area, or volume treated. The former is specified by the user when describing an application method for a EUPE (See Chapter 5), while the latter is a characteristic of the specific residence and the room (or pet) where the application occurs. The following calculations yield applied mass of AI for the application methods and MEs addressed by Version 2.0.

For an indoor fogger:

$$\text{AM} = \text{Room volume} * \text{Application Rate in } mg/m^3$$

For a broadcast by any method of application (indoor and out)

$$\text{AM} = \text{Floor area of room (or size of lawn or size of garden)} * \text{Application Rate in } mg/m^2$$

Note: In the case of the size of the lawn the area may be capped for products that are intended to treat small areas of the lawn. The user enters the maximum area treated in the AIPD.

For Crack and Crevice:

AM = Room perimeter * Application Rate in mg/m

For spot treatments indoor or on turf:

AM = Floor area of room (or size of lawn) * Fraction treated * Application Rate in mg/m^2

For treatments in the ornamental and "outdoors non-turf" MEs:

AM = Application Rate in mg * Area Treated

Because the size of the "outdoors non-turf" and ornamental MEs areas are defined. The user is asked to enter information on the amount of product used in an application. These data are then used to calculate the value of MA for an outdoor ornamental, and non-turf application.

For a bomb and pest strip, there is no applicator exposure, so the calculation is not made.

6.3.4 PROBABILITY OF HANDLER EXPOSURE FOR THE MODELED INDIVIDUAL

Handler exposures occur when the modeled individual applies the pesticide. It is possible for an individual to receive post-application exposures and not have been the handler. Therefore, the software determines whether the individual is the applicator of each pesticide used in her residence (home and yard).

The probability of the modeled individual being the applicator is determined in the following manner. Based on the findings in the NHGPUS (EPA, 1992), the majority of the applicators inside of homes are adult women, while the majority of applicators out of doors are adult men. Based on the data presented in Table 2.31 of the NHGPUS report, approximately 70% of the applicators were female for indoor application of pesticides

and approximately 70% of the applicators were male for outdoor application of pesticides. Following EPA policy, the model assumes that children living with their parents do not apply pesticides.

If the individual modeled is a female adult (ages 18 and older) in a residence where she is the head of the household[76], that there is a 70% (or alternative user-specified value) probability that she will be the applicator for indoor application of pesticides. A male adult (ages 18 and older) in a residence where she is the head of the household will have a 30% chance of being the applicator when pesticides are used indoors. The percentiles are reversed for out of door pesticide applications. Applicator exposure for the modeled individual is determined for each modeled application event.

6.4 Temporal Analysis of Exposure

6.4.1 BACKGROUND

A key feature that distinguishes Microexposure Event Analysis (MEA) in general, and LifeLine™ in particular, from earlier approaches to exposure assessment is that it assesses exposures to each modeled individual in a population on a series of successive days. Historical exposure assessment practice, reflecting both limited computer power and a corresponding inability to address available data sets, has in contrast evaluated exposures under fixed conditions, whether those conditions were expected to represent a single day or thirty years of exposure.

As a consequence of the use of such static exposure models, exposure assessments (unlike real exposures) were fragmented into either chronic assessments (using assumptions that represented average conditions over prolonged periods) or acute assessments (representing the extreme conditions that might be observed). It was not possible to evaluate the extent of variability over time in chronic assessments, nor to assess the frequency with which the conditions addressed in acute assessments might

occur. Accordingly, the wealth of information on how exposure duration influences toxic responses had to be distilled into a very limited number of arbitrary periods, ignoring biology in favor of analytical convenience.

The failure of such static exposure models to address probabilistic exposure assessment in a reasonable manner was, in fact, a driving motivation for the development of MEA (Price, 1996). The application of Monte Carlo techniques in static models has been particularly problematic for chronic exposure assessments: addressing the wide degree of variation in exposure that could occur over thirty years, for example, in a static Monte Carlo model is almost certain to yield unreasonable combinations of parameter values. While this issue presents a smaller problem for acute exposure assessments, historical practice has left the assessor unable to address the probability of sequential exposures. A range of acute exposures could be predicted, but it was not possible to say how likely it was that any individual would experience the modeled exposures from day to day.

By explicitly modeling individuals, rather than populations, and by using transition rules to ensure that each individual was modeled in a coherent, plausible, manner, MEA allows the exposure and risk assessor to evaluate dynamic conditions, rather than forcing exposure into a static model of fixed duration. Given adequate data and computer power, a single model can address exposures from seconds to an entire lifetime.

6.4.2 THREE MEASURES OF EXPOSURE

LifeLine™ determines the exposure history for each individual. This history consists of the route and source specific doses that occur on each day of the person's exposure simulation. When modeling more than a few individuals this results in a very large data file. The size of the files presents technical problems in terms of storage and access. In order to keep the output files a reasonable size. LifeLine™ Version 2.0 saves only a portion of the exposure data calculated. Three types of data are selected:

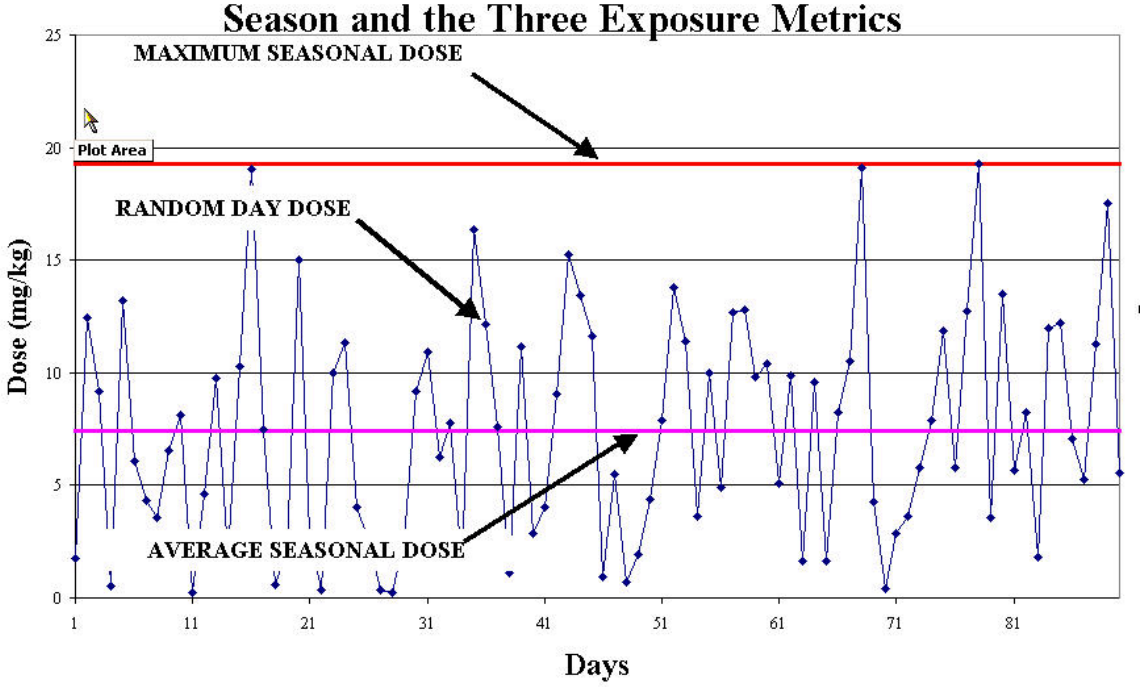
---

[76] An individual aged 18 or older is assumed to be the head of the household (and not a child living with their parents) if the have moved at anytime after there 18th birthday,

1. Average of the 90-92 daily doses that occur in a season for each of the three route-specific and three source-specific doses (average day).
2. The three route-specific and three source-specific doses daily doses that occur on a randomly selected day from the 90-92 days in the season (random day).
3. The highest of the 90-92 daily doses that occur by the three route-specific doses, the three source specific doses, and the total systemic dose (maximum day).

Figure 6-1 presents the three types (or metrics) of exposure graphically for one route. Note the random day metric may under or over estimate the average does in a season but will always be the same or lower than the maximum day.

Figure. 6-1. Daily Doses for One Individual over a Season and the Three Exposure Metrics



LifeLine™ Version 2.0 uses the data on each of the four seasons to derive the equivalent measures for an entire year. The determination of the annual values is calculated in the following manner.

1. The average annual dose is the average of the four seasonal average doses.

2. The annual random day dose is randomly selected from the four seasonal random day doses.
3. The maximum annual dose is the highest of the four seasonal maximum doses.

The random day metric provides a measure of the exposure that will occur on a randomly selected day. In analyses where there is no difference in the exposure potential from one portion of a season to another the random day give a measure of the distribution of doses that occur on a single day. Such a measure can be thought of as a snapshot of doses in the modeled population. Thus, if the random day distribution indicates that 95 % of the simulated 3 year olds have doses less then 20 ug/kg during spring. Then on any given day 95% of the nation's three year olds will have doses that are less than 20 ug/kg.

The seasonal and annual maximum doses present a measure of the upper bound of the interindividual variation in doses. Thus a finding that 95% of the population of three year olds have maximum annual daily doses less than 20 ug/kg, suggests that in the modeled population 95% of three year olds will not see doses in excess 20 ug/kg at any time between their third and fourth birthdays.

6.4.3 DAILY TIME-STEP AND EXPOSURE DURATION

6.4.3.1 Determining doses for periods of time greater than one day

Because exposures on successive days are modeled coherently for each individual in LifeLine™ the assessor is not limited to looking at single-day exposures. This distinguishes the system from models that address random days whether for discrete individuals or entire populations.

This design allows the user to address a "rolling average" (or alternately, a "rolling maximum" or other metric) for a period of any length between 1 and 365 days. Thus, for example, a user interested in an exposure period of 28 days could address the average exposure on days 1-28, 2-29, 3-30, and so on. The user need not be constrained to evaluating any particular 28-day period or, worse, to assuming constant conditions for a 28-day period. This feature allows the risk assessor to conduct an exposure assessment that is appropriate to the available toxicology data, rather than arbitrarily constraining the

toxicology to fit some predetermined acute or chronic exposure assumptions. As data increasingly become available on the temporal dependencies between exposure and biological effect, this capability will grow in importance.

In addition to the default daily exposure estimates, LifeLine™ allows the user to select as many as four additional exposure periods to be addressed in the analysis. On each modeled day, the system evaluates not only that day's exposure, but also calculates exposure for each selected period of n+1 days, using the new day's values in a weighted average with the previous n days of exposure. A typical use of the system would be to evaluate 1-day, 7-day, 28-day, 90-day, and annual average exposures.

As with the single day measure of exposure LifeLine does not save all of the averages. Rather the software saves the exposure metrics that parallel the single day metrics:

1 Average of the n-day average doses that occur on the periods ending on one of the days of the season (average n-day average).

2 The n-day average dose for the period that ends on a randomly selected day during the season (random n-day average).

3 The highest n-day average doses that occur on any period ending on a day of a season (maximum n-day average).

Finally, the model will always track the average daily dose over the entire period of the individual's life. These lifetime-average-daily doses can be used in the evaluation of carcinogenic risks. Thus, LifeLine™ is the first model that will allow the user to determine the distribution carcinogenic risks in the U.S. population associated with the aggregate exposure to AIs.

6.4.3.2 Additional uses of Variable Exposure Periods

While a primary use of user-defined exposure periods is to allow the exposure analysis to be tailored to toxicological data, collecting these data also facilitates a rapid summary of the predicted temporal patterns of exposure, without detailed examination of daily exposure records.

For example, if exposures are dominated by infrequent contact with high concentrations (e.g., an infrequently used residential product, or an upper-percentile food residue that differs significantly from the mean residue value), the assessor would likely find a significant difference between 1-day and 30-day exposure averages, but not between 30-day and annual averages. In contrast, if the driving exposure factor was relatively stable within a season, but showed marked change across seasons, there would be little difference between 1- and 30-day averages, but significant changes between 30-day and annual averages.

6.4.3.3 Periods of Time that are Less than One Day

As described earlier in this chapter, exposure events that occur over periods less than a day are integrated to yield a daily exposure estimate as the fundamental element of the system.[77] This is not an inherent characteristic of MEA model design, but rather a practical choice to accommodate the exposures of greatest interest without excessive demands on computer power. Future versions of LifeLine™ may include the determination of doses for periods of time that are shorter than one day. Such work would build on the "time of day" data in the NHAPS and CSFII.

6.4.4 LONGITUDINAL PATTERNS

A major gap in the available survey data is the lack of datasets that describe the behavior (dietary or activity) of large numbers of individuals over prolonged periods of time. While a few key studies extend observations over weeks, most of the available data collect data for only a few days in any individual.

In order to address this absence of longitudinal data, LifeLine™ allows the user to bound the impact of this data gap. The user is offered the choice of either assigning all of the variability in CSFII and NHAPS records to intra-individual changes over time, or to assigning all of that variability to differences between individuals. Specifically, the user can choose either to use a single record for an entire season, or to pick a new record for an individual from the appropriate bin on each day of a season. In the first case, the

individual's behavior (diet or activity pattern)[78] is assumed not to vary over time, all the variability in the sampled data is assumed to represent differences between individuals. In the second case, an individual is assumed to have no consistent behavior over time (beyond the characteristics that define a sampling bin), all of the variability within a bin is assumed to reflect intraindividual variability.

Providing this option supplies the exposure and risk assessor with a "brute force" solution to the confounding of intra- and interindividual variability in the key data sets. To the extent that the alternatives yield similar predictions of exposure or risk, no further investigation will be needed. Significant differences in model predictions suggest a need to develop data sets that overcome the inherent confounding in the principal data sources currently available.

6.5 References

Andelman, J. 1990 Total Exposure to Volatile Organic Compounds in Potable Water, Chapter 20 in Ram, N. et al. Significance and Treatment of Volatile Organic Compounds in Water Supplies, Chelsea, MI: Lewis Publishers, 485-504.

Giardino, N., E. Gumerman, J. Andelman, C. Wilkes, M. Small, J. Borrazo, & C. Davidson. 1990 Real-Time Measurements of Trichloroethylene in Domestic Bathrooms using Contaminated Water. *Proceedings of the Fifth International Conference on Indoor Air Quality and Climate*, Toronto, 2:707-712.

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.

US EPA. Dermal Exposure Assessment: Principles and Applications, Office of Health and Environmental Assessment, January 1992, EPA/600/8-90/011F).

---

[77] As noted elsewhere, the system routinely tracks seven measures of exposure on each day: three routes (oral, dermal, inhalation), three sources (food, tapwater, residential uses) and total exposure.

[78] Weekday and weekend activity patterns are always distinguished for non-dietary activities. See Chapter IV.

US EPA (US Environmental Protection Agency). 1992a. Guidelines for Exposure Assessment. Fed. Reg. 57(104)

US EPA, 1992b. *National Home and Garden Pesticide use Survey*. Prepared by the Research Triangle Institute for the Office of Pesticides and Toxic Substances, Biological and Economic Analysis Branch.

US EPA (US Environmental Protection Agency). 1997. Exposure Factors Handbook. EPA/600/P-95/002F(a-c), Washington, DC.

US EPA. 1998. Draft Standard Operating Procedures (SOPs) for Residential Exposure Assessments.

CHAPTER 7. MODELING RISK

While the conceptual framework for all toxic chemical risk evaluation is similar:

$$Risk = f(Exposure, Hazard)$$

there are a wide variety of metrics that are used to quantify this observation in different contexts. Perhaps the most familiar are the distinct approaches for assessing the risks of carcinogenesis (where risk is calculated as an estimate of the probability of an exposure-induced case of cancer), and for non-cancer toxic effects, where most practices yield an indication of whether an exposure is above or below a "safe" level[79]. Even within these broad categories, however, a wide variety of risk metrics has been developed, reflecting historical concerns in various assessment contexts.

7.1 Risk Characterization in LifeLine™

The fundamental goal of LifeLine™ is to be able to appropriately characterize the risks that are associated with the variety of estimates of exposure[80] that can be produced by the model, using toxicological endpoints of interest to the user. LifeLine™'s ability to produce a wide range of exposure and dose estimates (e.g., different routes and averaging periods) can be used to evaluate toxicity endpoints that range from acute non-cancer endpoints associated with a single day's exposure by a specific route to cancer risks associated with lifetime exposures.

The field of risk characterization is in flux. New methodologies are being developed for characterizing both cancer and non-cancer endpoints (e.g., EPA's draft cancer guidelines [EPA, 1996], and ILSI, 1999). Multiple approaches have been proposed for

[79] Some metrics give an indication of how far above or below the "safe" level an exposure is, although there is substantial debate over the toxicological meaning of such metrics.

[80] Depending upon user inputs, LifeLine™ determines exposure (the amount of an AI that reaches an individual) or the dose (the amount if AI that is absorbed), as well as risk. The software assumes that if risk calculations are being performed then the user has entered the absorption factors required to produce dose estimates appropriate to the toxicity data that have been entered.

characterization of non-cancer and cancer risks and regulatory policies for the acceptance or interpretation of these new methodologies have not been finalized.

Given the ongoing changes in assessment practice, it is not possible to design software that can capture all potential risk characterization methodologies. Version 2.0 of LifeLine™ has been designed to address those metrics of risk that are commonly used in the assessment of AIs. These include both carcinogenic risk expressed as a probability based on the cancer slope factor (also referred to as the q_1^*) as applied to Lifetime Average Daily Dose (LADD), and two comparisons of exposures/doses to data on non-cancer toxic hazards:

- Margins of Exposure (MOEs); and
- Fractions of the Reference Dose (RfD).

This decision makes it possible to produce an immediately usable system for aggregate and cumulative exposure and risk assessment[81], while focusing the first round of scientific dialog in the user community on the vastly expanded capabilities of exposure assessment offered by the system. The potential use of these capabilities to support more complex risk characterization techniques is deferred to later versions of the software.

7.1.1 TEMPORAL CONSIDERATIONS

While allowing the analyst to replicate historical indices of risk, the analysis of risk in LifeLine™ is also designed to reflect the realism and flexibility of the exposure modeling provided by the system. While LifeLine™ easily supports the modeling of both exposures and risks in terms of "acute," "sub-chronic," and "chronic" categories, it expands this capability to allow the user to tailor the duration of the exposure metric to the to the duration of dosing required to product a specific toxic effect for an AI.

Accordingly, rather than restrict the user to pre-defined categorization of toxicity data, LifeLine™ allows her to specify the relevant exposure durations for any toxicity metric. For example if a NOAEL is observed in a 28 day study, the user may wish to apply the

NOAEL to the evaluation of durations of exposure from 1-28 days. The software can specify that this NOAEL can be applied to doses of these durations. In each model run, the outputs of a specified duration will be matched to the toxicity data that the user has indicated as being comparable. This ability frees the risk assessor from the arbitrary categories of "acute" or "chronic," and allows the user either to match the exposure assessment directly to available toxicity data.

7.1.2 RISK AND THE EXPOSURE METRICS

As discussed in the prior chapter, LifeLine™ Version 2.0 LifeLine produces three types of exposure metric for each route of exposure and each averaging period. These measures are used to calculate the corresponding measures of risk characterization, MOE and percent RFD. LifeLine™ Version 2.0 produces three measures of the MOEs. For the one-day exposures, the three metrics are:

1 The total and route specific MOEs that are based on the average of the daily doses that occur in a season (average day MOE).

2 The total and route specific MOEs that are based on the daily doses that occur on a randomly selected day during the season (random day MOE).

3 The total and route specific MOEs that are based on the highest dose that occurs on any given day of a season (maximum day MOE).

Because the term "maximum day" in maximum day MOE refers to the dose used to derive the MOE, this measure of the individual's risks actually has the lowest value for any day of the season.

The determination of the maximum day MOE is determined independently for each route and for the total MOE. Thus, the maximum MOE for inhalation may come from one day during a season and the maximum MOE from dermal from another day. However, the total MOE is always based on the route specific doses that occur to an individual on the same day.

---

[81] If one is using a Toxic Equivalency Factor approach to assess multiple AIs, Version 2.0 also supports cumulative risk assessments.

The values of the MOEs for periods longer than a single day are based on the average doses for the selected periods. Thus the 7-day average maximum day total MOE is the lowest total MOE that occurs in a season as the result of the averages of the route-specific doses taken from a seven consecutive day period during that season.

LifeLine™ Version 2.0 uses the data on each of the four seasons to derive the equivalent measures for an entire year. The determination of the annual values is calculated in the following manner.

1 The average annual MOE is the MOE based on the average of the four seasonal average doses.

2 The annual random day is randomly selected from the four seasonal random day MOEs.

3 The maximum annual MOE is the lowest of the four seasonal maximum MOEs.

A similar approach is used for the calculation of the percent RFD.

7.1.3 RISK BY ROUTE(S) OF EXPOSURE

LifeLine™ allows the user to independently address the risks of route-specific and systemic toxic effects. For a substantial number of AIs, there are not only quantitative differences in systemic toxicity attributable to route of exposure, but also qualitative differences in the patterns of toxicity seen following exposure by different routes.

As discussed in Chapter 6, the system does produce estimates of route-specific exposures or doses. These route-specific doses are used to evaluate risk as the consequence of particular routes of exposure, in addition to the evaluation of aggregate risk by all routes of exposure. One of the interesting implications is that certain substances may be of greater concern for route-specific effects than for aggregate systemic risks,

7.1.3.1 Calculating Risks for Aggregate Exposures

7.1.3.1.1 Aggregation of Multipathway Risks

While the system develops estimates of route-specific risks, its major focus is the determination of aggregate exposure and the corresponding risk. In recent historical practice, the term "aggregate risk" has been used to refer to many different types of risk calculations. Version 2.0 of LifeLine™ determines aggregate risk using two approaches, route specific and systemic. The route-specific approach uses the route-specific doses and NOAELs. The total daily MOE is determined at the end of each day using the following formula:

$$\text{Total MOE} = \frac{1}{((\text{Oral NOAEL}/\text{Oral Dose})+(\text{Dermal NOAEL}/\text{Dermal Dose})+(\text{Inhalation NOAEL}/\text{Inhalation Dose}))}$$

If the NOAELs on any of the routes have not been entered in the *AIPD*, then a default value of zero should be used. If on a given day one or more of the route-specific doses is zero then the portion of the equation dealing that route should be dropped out. (Otherwise the equation will divide by zero and the calculation will give an error message.) Thus if the dose for inhalation is zero then the equation for total MOE will reduce to:

$$\text{Total MOE} = \frac{1}{((\text{Oral NOAEL}/\text{Oral Dose})+(\text{Dermal NOAEL}/\text{Dermal Dose}))}$$

If there is no dose for any of the routes on a given day, then the value of Total MOE should be set at zero.

The second approach, systemic is based on the estimate of the systemic (absorbed) dose that occurs as the result of multiple routes of exposure to one individual on one day or on a series of days. Accordingly, to assess risks from such aggregate exposure, the user is required to enter toxicity data that are relevant to systemic effects.

The concept of defining systemic NOAEL is not commonly required in current pesticide assessments. The systemic NOAEL can be thought of as the NOAEL associated with non-portal effects. The systemic NOAEL will be similar to the oral dose for compounds that are well absorbed by the oral route and are not subject to a first-pass effect.

7.1.3.2 Source-Specific Risks

The software also evaluates source-specific exposures and doses. Source-specific doses for residential and tapwater sources (and other sources to be addressed by later versions) can occur by multiple routes. Accordingly, risk estimates for these residential and tapwater doses are based on systemic effects. Doses from diet, in contrast, are evaluated using oral toxicity data. Tables 7-1 and 7-2 below summarize the toxicity information used for each type of risk estimate.

Table 7-1. Toxicity Data used in Non-Cancer Risk Characterizations

Exposure Estimates	Absorption Factor[82]	Toxicity Data
Oral Exposure	Oral Absorption	Oral toxicity data[83]
Inhalation Exposure	Inhalation Absorption	Inhalation toxicity data
Dermal Exposure	Dermal Absorption	Dermal toxicity data
Dietary Exposure	Oral Absorption	Oral toxicity data
Residential Exposure	Multiple Routes	Systemic or route-specific toxicity data
Tapwater Exposure	Multiple Routes	Systemic or route-specific toxicity data
Aggregate Exposure	Multiple Routes	Systemic or route-specific toxicity data

[82] The absorption factor must be applied appropriately. It would not be appropriate, for example, to incorporate an oral absorption factor into an exposure assessment for oral exposures, if the relevant toxicity data reflected administered dose by the oral route.

[83] e.g., NOAEL, RfD

Table 7-2. Toxicity Data used in Cancer Risk Characterizations		
Exposure Estimates	**Absorption Factor**	**Toxicity Data**
Lifetime Average Oral Dose	Oral	Oral potency data
Lifetime Average Inhalation Dose	Inhalation	Inhalation potency data
Lifetime Average Dermal Dose	Dermal	Dermal potency data
Lifetime Average Dietary Dose	Oral	Oral potency data
Lifetime Average Residential Dose	Multiple Routes	Systemic potency data
Lifetime Average Tapwater Dose	Multiple Routes	Systemic potency data
Lifetime Average Aggregate Dose	Multiple Routes	Systemic potency data

Whether a particular characterization of risk can be performed by the system is determined by whether or not the user has supplied relevant toxicological information.

7.1.4 FLEXIBLE INCORPORATION OF TOXICITY DATA

A goal of Version 2.0 of the software is to allow the use of a wide variety of types of toxicity data as the basis for non-cancer risk characterizations including, LOAELs, NOAELs, benchmark doses, and estimates of ED_{10}. The system makes any assumptions that the user has made concerning the toxicity data transparent. This includes not only the basic toxicological data employed, but also uncertainty factors, modifying factors, and the use (if any) of an additional factor to generate a Population-Adjusted Dose under FQPA.

7.1.5 CUMULATIVE RISKS

Version 2.0 of LifeLine™ can also be used to address cumulative risks where cumulative risks are defined in terms of the cumulative dose to an index AI. This is known as a RPF (Relative Potency Factor) or TEQ (Toxic Equivalent) approach. In this type of analysis,

[84] In conventional risk assessment practice, the same data set may be used to provide an estimate of both systemic and route-specific toxicity. It is quite common, for example, to use toxicological data from oral exposures in the prediction of systemic effects as well as route-specific effects.

the exposure metrics for each AI being evaluated are converted into equivalents (TEQs) amounts of the index AI.

For AIs A and B, where the RfD for A is twice that of B and A is selected as the indicator AI, the adjustment would be as follows:

Dose A' = Dose B x 2

Where:

Dose B is the actual exposure metric for AI B.
Dose A' is the equivalent exposure metric for the indicator AI A.

In the case of dietary exposures, data on the co-occurrence of the residues of the pesticides in a food are converted to TEQs for an index AI and summed to give a distribution of TEQs for that food. These distributions of TEQs can be entered into LifeLine™ as if they were the concentrations of a single compound. A similar approach is used for the evaluation of tapwater exposures. Data on concurrent levels of AIs measured in surveys of water supplies are converted to a single distribution of TEQ and entered into LifeLine™.

A somewhat different approach is used in the assessment of residential sources of exposures. For these sources of exposure, the amounts of each of the AIs applied during the use of a specific product are converted to the corresponding TEQs. Then LifeLine™ is run with all of the products that contain any of the compounds.

The toxicity data for the index AI are entered into the model to generate risk estimates. The result is an estimate of the cumulative risks associated with the cumulative exposures to all of the AIs. Any of the risk characterization metrics can be used to evaluate the cumulative risks.

7.2 User Specification of Toxic Hazard Data

7.2.1 TOXICITY DATA: IDENTIFYING THE TYPES OF RISK TO BE EVALUATED

The user defines the range of risk estimates that can be performed by the system by supplying the corresponding toxicity data. Data on the toxicology of an AI are entered in the *Active Ingredient and Product Description* program (under *Enter or Edit Model Inputs*). The user begins by selecting up to three different sets of non-cancer toxicity data (distinguished by relevant exposure durations) and indicating whether or not to evaluate carcinogenic risk.

The three types of non-cancer toxicity data are distinguished by the exposure durations for which they are relevant. For convenience, they are labeled *short-term*, *intermediate-term*, and *long-term*, but each can address any exposure duration between one day and a year.[85] For each selected type of toxicity data, the user specifies the minimum and maximum relevant duration of exposure. Cancer is always evaluated based on Lifetime Average Daily Dose.

The selected toxicity types cannot be overlapping. Thus, for example, if the user has selected *short-term* and *long-term* toxicity, the minimum duration for long-term toxicity must exceed the maximum duration for *short-term* toxicity. Similarly, if all three types are selected, they must represent three distinct sets of durations. (The system will not allow the user to specify overlapping durations.)

7.2.2 ROUTE-SPECIFIC AND SYSTEMIC TOXICITY DATA

Both cancer and non-cancer risk characterizations require the matching of toxicological criteria to the appropriate route of exposure, as described above. Accordingly, for each type of toxicity data selected by the user, separate values of the relevant toxicity parameters can be specified for each of the three routes of exposure (oral, dermal, inhalation) and for systemic toxicity. The data entry screens facilitate copying toxicity

data among routes (for example, where oral toxicity data are used to predict systemic toxicity and/or toxicity by other routes of exposure)

The route designations for entering toxicity data do **not** reflect the route of administration used in the underlying toxicology study, but rather address the **applicability** of the toxicology data to human health risk assessment. Data collected from a study using any route of exposure may be applicable to the assessment of toxic effects on a systemic basis, even when exposure occurs primarily by other routes. If the user wishes to evaluate risks that reflect exposures from multiple routes of exposure, she **must** supply a corresponding data set for **systemic** toxicity. Route-specific hazard data are used solely to evaluate risk as the consequence of particular routes of exposure.

7.2.3 TOXICITY PARAMETERS - NON-CANCER

For each route of exposure, as well as for systemic toxicity, the user is able to specify between one and four toxicity parameters. At a minimum, the user must specify one toxicity parameter for at least one route of exposure (or for systemic toxicity) for each selected type of non-cancer toxicity. The system uses the data entered by the user to calculate two additional parameters.

The four parameters that may be specified by the user are:

- A **Toxicity Measure** in milligrams per kilogram body weight per day, such as:
  - ⇒ No Observed Adverse Effect Level (NOAEL),
  - ⇒ Lowest Observed Adverse Effect Level (LOAEL),
  - ⇒ Dose producing an effect in 10 percent of exposed subjects (ED_{10}), or
  - ⇒ Benchmark Dose,[86]
- An **Uncertainty Factor** (UF, unitless) that reflects differences between the design of the study and likely exposures to humans,

---

[85] Multi-year exposures are addressed on an annual basis.

[86] The user is able to specify the nature of the toxicity measure used for each exposure route and for systemic toxicity.

- A **Modifying Factor** (MF, unitless) reflecting confidence in the applicability of the study to human risk, and
- An **FQPA factor** (unitless), reflecting special toxicity concerns applicable to children or women of childbearing age.

Only the first of these is required (for at least one route or for systemic). The user is free to specify each value independently for any route of exposure. Thus, for example, the same NOAEL could be applied to oral and inhalation routes, but different uncertainty, modifying, or FQPA factors could apply to each.

Because the **FQPA Factor** used to determine the population-adjusted dose generally reflects developmental toxicity or the special sensitivity of children to the toxic effects of the AI, the user has the option of specifying the maximum age at which this factor will apply, both for route-specific and systemic effects. Because there may be concerns about differences between the sexes, or concerns about effects from prenatal exposures, separate ages may be specified for males and females. The defaults provided are intended to reflect males as children and both female children and females of childbearing age.

If the user has specified, in addition to the **toxicity measure**, an **uncertainty facto*r*** and a **modifying factor** for any route of exposure, the system will calculate corresponding Reference Dose (RfD) applicable to general populations [RfD = NOAEL / (UF * MF)]. If she or he has also specified an ***FQPA factor*****,** the system will also calculate a Population-Adjusted Dose (PAD) reflecting special concern for risks to children or from prenatal exposure [PAD = RfD / FQPA factor].

7.2.4 CARCINOGENIC HAZARD DATA

If the user has elected to specify toxic hazard data for carcinogenicity, the system collects information on the "potency" of the AI as a carcinogen (i.e., how quickly risk increases with increases in dose), on the level of confidence that the AI is in fact carcinogenic in humans, and on the source of the assessment of carcinogenic hazard. Only the data on the potency is used in the quantitative risk assessment. Data on the classification of the

AI as a carcinogen are retained and made available in the report for the risk group that is generated by the *Active Ingredient and Product Description Program*.

As with the data on non-cancer effects, the software allows the user to define separate data for the evaluation of route-specific and systemic doses. As in the case of non-cancer assessments, the potency for systemic doses is used in the evaluation of risk from aggregate exposure.

The rate of increase in risk with increasing exposure is expressed as the slope of a function that is presumed to be linear at low doses.[87] This Slope Factor is sometimes referred to as a q_1^* or q^*.

There are two parallel rating systems used to designate scientific confidence that a particular AI is carcinogenic in humans, based upon available data. The EPA uses letters as primary designations, with numbers to indicate subdivisions, while the International Agency for Research on Cancer (IARC) uses the inverse approach. The categories are quite similar, the following table presents their general meaning, although there are subtle terminological differences. The system allows the user to enter a classification using either system.

Table 7-3

Carcinogenicity Weight of Evidence Classification	**EPA**	**IARC**
Known human carcinogen	A	1
Probable human carcinogen – human data	B1	2A
Probable human carcinogen – animal data	B2	2B
Possible human carcinogen	C	3
Not classifiable as to carcinogenicity	D	4
Evidence that not a carcinogen in humans	E	5

The system also asks the user to specify whether the source of the assessment of human carcinogenicity was EPA, IARC, or another source.

In this version of the system, only the non-threshold (linearized) approach to estimating the probability of cancer from exposure is addressed. Recent research indicates that this may not be an appropriate risk model for all carcinogens, just as threshold models appear to be inadequate for addressing the non-cancer toxic effects of some agents.

7.2.5 LOCATION OF THE TOXIC HAZARD DATA

Because data on toxic hazards represent a characteristic of the AI, they are collected along with other information on the AI, such as physical-chemical properties (in the *Active Ingredient and Product Description* program). Correspondingly, they are stored in the *Risk Group* file for the AI (or set of AIs).[88]

7.3 Matching Exposure Assessment to Toxicity Data

Risk estimates must be based on exposures that are appropriate to the available toxicity data. Therefore, the system provides information to guide the user in evaluating exposures for which she has toxicity data when she is setting the parameters for an exposure analysis. This guidance occurs in two places in the opening section of the *LifeLine Model*.

7.3.1 LIFESPAN

As noted in Chapter 1, LifeLine™ provides for the analysis of exposures on each day of an entire lifetime, for every individual in a defined population. As part of this analysis, the system evaluates lifespan for each individual based on demographic characteristics. When these mortality calculations are employed, modeled individuals die at every age

[87] This version does not address non-linear models of cancer risk, which may be more appropriate for certain AIs.

[88] Each exposure or risk analysis in LifeLine™ requires a *risk group* file that describes one or more AIs.

from 1 year to 85 years (Available mortality data do not allow an accurate model of the distribution of life spans for individuals who complete their 85th year).

Such a collection of varied life spans presents a major challenge for the traditional assessment of cancer risk that is based on Lifetime Average Daily Dose. In particular, individuals who die in childhood have essentially no risk for many exposure-induced cancers, because they do not survive the latency period of the cancer. Moreover, the historical calculation of LADD has been based upon the assumption of a fixed lifespan (generally 70 or 75 years).

There are a number of ways around this difficulty, as well as enhancements (for example, to address the ages at which exposures occur) that will be explored in future versions of the system. Version 2.0 of LifeLine™ uses the simple expedient of requiring the user to specify a fixed lifespan of at least 70 years when evaluating cancer risks. It then calculates out the actual lifetime average daily dose for use in cancer risk assessments. For populations in which mortality is addressed, LADD but not cancer risk can be evaluated. This is addressed in the "**General**" tab of **Analysis Preferences** in the *LifeLine Model.*

7.3.2 EXPOSURE PERIODS

As noted above, a key feature provided by the system is the ability to match the time course of toxicology and exposure without resorting to arbitrary categories such as "acute," and "chronic," and all toxicology data are entered with an indication of the exposure durations over which they are applicable.

The matching of exposure analysis to toxicology data is addressed in the "**Averaging Period / Toxic Hazard**" tab of **Analysis Preferences** in the *LifeLine Model*. When the user selects exposure periods to evaluate (in addition to the default evaluation of daily exposure), the system displays the toxic hazard data that are available in the selected *Risk Group* file, including the minimum and maximum exposure durations to which they apply, and the type of data available for each route of exposure and/or systemic toxicity.

7.4 Risk Characterization

7.4.1 CHARACTERIZING NON-CANCER RISKS

Non-cancer risks for dietary exposure have historically been estimated as a fraction of the RfD. Non-cancer risks for residential have been assessed in terms of a margin of exposure (MOE) model. LifeLine™ supports both approaches for both route-specific and systemic risks.

The system has the following capabilities:

- Calculation of route-specific risk estimates, and **where toxicologically appropriate data are available**, aggregate- and source-specific estimates of risk;
- Separate characterization of risk for different time periods using independent hazard measures;
- Maximizing flexibility by allowing the user to assign specific exposure periods for evaluation within these categories;
- Allowing the user to analyze exposure or dose, and report dose, risk, or both when toxicity data have been specified;
- Separate calculation of cancer risk using a linear non-threshold potency estimate (q_1^*) applied to the lifetime average daily dose (LADD); and
- Calculation of non-cancer risk both as percent RfD and as a MOE, for all routes of exposure.

Application of the toxicology data that were entered in the *Active Ingredient and Product Description* program to characterize risk is performed in the *Report Generator and Viewer* program. This approach provides for computational efficiency.

7.4.1.1 Rules for Data Use

As noted above, there may be up to four sets of data on non-cancer toxic hazards (oral, dermal, inhalation, and systemic) for each of three exposure periods. There may also be four sets of data on cancer hazard. The risk evaluations that are supported by the system depend upon which of these sixteen data sets are available. For non-cancer risk

characterization, the range of possible evaluations also depends upon how complete each dataset is.

MOE Calculations - For each MOE calculation, the user must have specified a basic toxicological datum (NOAEL, etc.) for the appropriate route (or systemic, if aggregate risk or source-specific risk for non-food sources).

Percent RfD Calculation - For each percent RfD calculation, the user must have specified, in addition to the basic toxicological datum (NOAEL, etc.), an Uncertainty Factor and Modifying Factor. FQPA factors will be used if specified, but are not mandatory. If, however, an FQPA factor has been specified, the user must specify that ages of males and females where the factor is applied.

Cancer Risk Calculation - Slope data are required for the risk calculation. The system automatically collects data on WOE class and the source of the carcinogenic assessment.

7.4.1.2 Calculation of the MOE

Then general form of the Margin of Exposure (MOE) is to divide a NOAEL by the predicted exposure:

MOE = NOAEL / Dose

Where the route-specific absorption factor has been set at one, dose is equivalent to exposure. In LifeLine™, **any** user-supplied *Toxicity Measure* can be used to calculate an MOE. This requires that the user who has specified a value such as a benchmark dose be aware that the resulting MOE may have a different meaning than the traditional value. The *Toxicity Measure* (e.g. NOAEL) **for the appropriate duration and route** is divided by the corresponding route-specific exposure currently reported in the output. The following are the doses that are matched to the specific MOE estimates.

7.4.1.2.1 Aggregate

Depending on the method chosen aggregate MOEs can be calculated by either the route specific toxicity measures or the ***systemic*** *Toxicity Measure* (e.g., NOAEL) for the appropriate duration.

7.4.1.2.2 Source-Specific (Residence and Tapwater)

The ***systemic*** *Toxicity Measure* (e.g., NOAEL) for the appropriate duration is divided by the corresponding source-specific aggregate dose.

7.4.1.2.3 Source-Specific (Food)

The ***oral*** *Toxicity Measure* (e.g. NOAEL) for the appropriate duration is divided by the corresponding dietary aggregate dose.

7.4.1.2.4 Route-Specific

The ***oral***, ***inhalation***, or ***dermal*** *Toxicity Measure* (e.g. NOAEL) for the appropriate duration is divided by the corresponding route-specific dose.

7.4.1.3 Calculation of Percent RfD

The general form of this calculation is as follows:

Percent RfD = Exposure (dose) / RfD * 100

Dividing the *Percent RfD* by 100 yields another typical index of non-cancer toxicity, the *Hazard Quotient*.

7.4.1.3.1 RfD and PAD

The use of an FQPA factor to calculate a Population Adjusted Dose (PAD) adds an additional factor in the calculation of the Percent RfD. As noted above, a PAD may only be applicable within a certain age range, and the relevant age range may be different for males and females.

If, for example, an FQPA factor of 10 applies to males up to 12 years of age and to females up to 45 years of age, the following behavior will be seen in the Percent RfD in the male and female populations. Assuming, that exposure is held constant,[89] at age 12, the Percent RfD values of all the males will decrease by a factor of 10, at age 45, the

values for females will decrease by a factor of 10. Accordingly, the user who is generating Percent RfD measures must remain aware of the ages and sexes of the individuals for which the value is reported. In the above example, for a population of 20-year-olds the males will not have the FQPA factor applied to them but the females will. Thus, the population average is a mix of two different toxicity measures.

Because data on the sex of a modeled individual is stored in the LIVES.DBF log file, generating this file is a prerequisite for generating Percent RfD with an FQPA factor. An appropriate notice appears on the corresponding window.

The Percent RfD is determined for each route by dividing the dose ***for the appropriate duration and route*** by the RfD or PAD and multiplied by 100. The following are the doses that are matched to the specific Percent RfD estimates.

7.4.1.3.1.1 Aggregate

The dose is divided by the ***systemic*** RfD or PAD, and multiplied by 100.

7.4.1.3.1.2 Source-Specific (Residence and Tapwater)

The dose is divided by the ***systemic*** RfD or PAD, and multiplied by 100.

7.4.1.3.1.3 Source-Specific (Food)

The dose is divided by the ***oral*** RfD or PAD, and multiplied by 100.

7.4.1.3.1.4 Route-Specific

The oral, inhalation, and dermal doses are divided, respectively, by the **oral, inhalation, or dermal** RfD or PAD, and multiplied by 100.

7.4.2 CALCULATION OF CANCER RISK

As noted above, this calculation is only available when the exposed population was evaluated with a population that had a fixed lifespan of at least 70 years, as specified in *Analysis Preferences* in the *LifeLine Model*.

[89] This will rarely be observed in an actual exposure assessment.

First, the system calculates LADD for each route of exposure.[90] Risk is then calculated as the product of dose and slope factor using the following equation:

$$Risk = 1 - e^{-(slope\ factor * dose)}$$

Because this risk calculation applies only to LADD, there is no calculation of risk at specific ages. The model thus produces a single distribution of inter-individual variation in cancer risks for the modeled population.

The following are the LADDs and route-specific slope factors use in estimating the route and source specific carcinogenic risks.

7.4.2.1.1 Aggregate

The ***systemic*** LADD is evaluated against the ***systemic*** slope factor (q_1^*).

7.4.2.1.2 Source-Specific (Residence and Tapwater)

The ***source-specific*** LADD is evaluated against the ***systemic*** slope factor (q_1^*).

7.4.2.1.3 Source-Specific (Food)

The ***source-specific*** LADD is evaluated against the ***oral*** slope factor (q_1^*).

7.4.2.1.4 Route-Specific

The ***route-specific*** LADD is evaluated against the ***route-specific*** slope factor (q_1^*).

7.5 References

US EPA 1996. Proposed Guidelines for Carcinogen Risk Assessment, National Center for Environmental Assessment. EPA/600/P-92/003C, April 1996

ILSI, 1999 A Framework for Cumulative Risk Assessment, An ILSI Risk Science Institute Workshop Report, Ed. B. Mileson, E. Faustman, S. Olin, P. Ryan, S Ferenc, T Burke, Washington DC

[90] This calculation is available for populations that are evaluated with shorter lifespans (or with mortality evaluation), but is not applicable to the cancer risk calculation. The user must exercise care in interpreting such values.

US EPA 1991. Environmental Protection Agency National Oil and Hazardous Substances Pollution Contingency Plan under the Comprehensive Environmental Response Compensation, and Liability Act of 1980. US Environmental Protection Agency, Washington, DC. April 19 (1991)

US EPA, 2001, Gguidance on Aggregate Exposure Assessments.