The Stochastic Human Exposure and Dose Simulation Model for Multimedia, Multipathway Chemicals (SHEDS-Multimedia): Residential Module

SHEDS-Residential version 4

Technical Manual

May 21, 2012

Prepared by: Graham Glen[1], Valerie Zartarian[2], Luther Smith[1], and Jianping Xue[2]

[1] Alion Science and Technology, Inc.

[2] US Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory

Disclaimer

EPA's SHEDS-Residential version 4 model is a physically-based, probabilistic model that can simulate for a population of interest cumulative (multiple chemicals) or aggregate (single chemical) residential exposures over time via multiple routes of exposure for different types of chemicals and scenarios. The United States Environmental Protection Agency, through its Office of Research and Development, developed and funded SHEDS-Residential version 4 with assistance from contractor Alion Science and Technology, Inc.

SHEDS-Residential version 4 is one module (along with the separate SHEDS-Dietary module) of EPA's more comprehensive human exposure model, the Stochastic Human Exposure and Dose Simulation model for multimedia, multipathway chemicals (SHEDS-Multimedia), which can simulate aggregate or cumulative exposures over time via multiple routes of exposure (dietary & non-dietary) for different types of chemicals and scenarios. SHEDS-Residential and SHEDS-Dietary will be merged together in a future version of SHEDS-Multimedia.

SHEDS-Residential version 4 includes a case study example for illustrative purposes, as well as a default file for non-chemical specific inputs, as described in the Technical Manual and User Guide. All input values used in the SHEDS-Residential model for a given application should be entered or reviewed by the researcher so that the model results are based on appropriate data sources for the given application.

Version 4 of SHEDS-Residential reflects comments from EPA's August 2007 external Scientific Advisory Panel that reviewed SHEDS-Multimedia version 3 (the aggregate residential version). SHEDS-Residential version 4 will undergo external peer review by EPA's Scientific Advisory Panel July, 2010, and should be considered draft at this time.

Acknowledgments

We would like to thank David Miller of the U.S. Environmental Protection Agency, Office of Pesticide Programs and Andrew Geller of the U.S. Environmental Protection Agency, Office of Research and Development for providing comments on the draft Technical Manual.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND ABBREVIATIONS

ADD – average daily absorbed dose

CCA –Chromated Copper Arsenate

CDFs – cumulative distribution functions

CHAD – Consolidated Human Activity Database

DD – Daily Dose

EOHSI – Environmental Occupational Health Sciences Institute

EPA – United States Environmental Protection Agency

EPOC – post-exercise oxygen consumption

FIFRA – Federal Insecticide, Fungicide, Rodenticide Act

g – gram

GI – gastrointestinal

GM – geometric mean

GSD – geometric standard deviation

GUI – graphical user interface

LBNL – Lawrence Berkeley National Laboratory

Mets – metabolic ratio of energy expenditure for an activity to the resting rate

NERL – National Exposure Research Laboratory

NHANES – National Health and Nutrition Examination Survey

OPP – Office of Pesticide Programs

ORD –Office of Research and Development

PBPK – physiologically-based pharmacokinetic

PDFs – probability density functions

PK- pharmacokinetic

SA – surface area

SAP – (FIFRA) Scientific Advisory Panel

SHEDS – Stochastic Human Exposure and Dose Simulation

UPA – University Partnership Agreement

ug (in SAS printout or variable names) – microgram

yr – year

1 INTRODUCTION

1.1 SHEDS-Multimedia Overview

Reliable models for assessing human exposures are important for understanding health risks from chemicals. The Stochastic Human Exposure and Dose Simulation model for multimedia, multiroute/pathway chemicals (SHEDS-Multimedia) is being developed by the U.S. Environmental Protection Agency (EPA)'s Office of Research and Development (ORD), National Exposure Research Laboratory (NERL) as a state-of-science computer model for improving estimates of aggregate (single-chemical, multi-route/pathway) and cumulative (multi-chemical, multi-route/pathway) human exposure and dose. SHEDS-Multimedia is the EPA/ORD's principal model for simulating human exposures to a variety of multimedia, multipathway environmental chemicals such as pesticides, metals, and persistent bioaccumulative toxins. SHEDS-Multimedia is coded in SAS version 9.1; the user must have this version (or later) of SAS installed prior to the installation of SHEDS.

SHEDS-Multimedia version 4 is comprised of both a residential module (SHEDS-Residential version 4.0), described in this technical manual and a related user guide (Isaacs et al., 2010a), and a dietary module (SHEDS-Dietary version 1.0; Xue, 2010b; Xue et al., 2010c; Isaacs et al., 2010b). SHEDS-Dietary is a probabilistic, population-based dietary exposure assessment model that simulates individual exposures to chemicals in food and drinking water over different time periods (e.g., daily, yearly). It can produce population percentiles of dietary exposure by source and age-gender group; quantify contribution to total exposure by food, commodity, and chemical; and be used for eating occasion, sensitivity, and uncertainty analyses (Xue et al., 2010b). The dietary model can use either USDA's CSFII (1994-96, 1998) or NHANES/WWEIA dietary consumption data (1999-2006), along with EPA/USDA recipe translation files (Food Commodity Intake Database), and available food and water concentration data.

A methodology for linking the residential and dietary modules for simulated individuals (based on age, gender, body weight, total caloric intake/METS, race, season, weekday and region) will be peer reviewed by EPA's July 20-22, 2010 FIFRA Scientific Advisory Panel. This methodology has been tested through "soft linking" the two modules with a permethrin pesticide case study. In the next version of SHEDS-Residential, the residential and dietary module SAS codes will be merged, so that both types of exposure can be calculated for the same individual after food consumption and activity pattern diaries are appropriately matched. A common Graphical User Interface (GUI) will also allow the user to run either module separately, or to run them both together. The focus of this Technical Manual is the standalone SHEDS-Residential model.

SHEDS-Residential is a physically-based, probabilistic model that predicts, for user-specified population cohorts, exposures incurred in the residential environment over time via inhaling contaminated air, touching contaminated surface residues, and ingesting residues from hand- or object- to-mouth activities. To do this, it combines information on chemical usage, human activity data (e.g., from time/activity diary surveys, videography studies), environmental residues and concentrations, and exposure factors to generate time series of exposure for simulated individuals. One-stage or two-stage Monte Carlo simulation is used to produce distributions of exposure for various population cohorts

(e.g., age/gender groups) that reflect the variability and/or uncertainty in the input parameters (See Appendix A for a definition of these terms).

While the core of SHEDS-Residential is the concentration-to-exposure module, there are several options for obtaining concentration inputs (described in detail later), and SHEDS-Residential exposure outputs can be used as inputs to physically-based pharmacokinetic (PBPK) models. While SHEDS-Residential can simulate chemical application scenarios (residential indoor crack and crevice, pets, indoor fogger, lawn, and garden, for different application methods), it can also simulate exposures to chemicals already present in various environmental media.

This Technical Manual describes the algorithms, methodologies, and input and output capabilities of SHEDS-Residential version 4.0, which can simulate either cumulative (multi-chemical) or aggregate (single chemical) exposures from inhalation, dermal contact, and non-dietary ingestion routes. Version 3 focused on residential aggregate exposures only (Zartarian et al., 2008). This version 4 Technical Manual also includes some detail of the SHEDS-Multimedia residential module SAS code (Glen, 2010); details of the SHEDS-Multimedia dietary module (SHEDS-Dietary) are provided in a separate technical manual (Xue et al., 2010b) and user guide (Isaacs et al., 2010b). Reviewed collectively, the annotated SHEDS-Residential and SHEDS-Dietary SAS codes, Technical Manuals, graphical user interfaces (GUI), and User Guides should provide a thorough perspective and understanding of the overall SHEDS-Residential version 4.0 technical and usability aspects.

Other new features in SHEDS-Residential version 4.0 include cumulative algorithms (multiple chemicals and their co-occurrence in space and time); reproducibility of prior runs (in whole or in part); a new option for constructing longitudinal activity diaries; the option to use empirical distributions; exposure estimates for those who handle or apply the chemicals; calculations of margins of exposure (MOEs); and a new sensitivity analysis option (Glen et al., 2010). Also, SHEDS-Residential version 4.0 outputs files can be input to detailed physiologically-based pharmacokinetic models.

The overall SHEDS-Multimedia model, and its individual residential and dietary modules, can help answer many useful questions related to populations' cumulative or aggregate exposures for different multimedia chemicals, and what factors and pathways are most important. This model can be linked with other tools (e.g., fate and transport models, dose estimation models, measurements) for characterizing uncertainty in risk assessments. Anticipated primary end users of SHEDS-Residential include EPA Program Offices such as the Office of Pesticide Programs, other federal and state agencies, and the scientific community. Other SHEDS models, with similar approaches but different code and graphical user interfaces, have been developed by EPA/ORD to address exposures to particulate matter (SHEDS-PM; Burke and Vendatham, 2009), air toxics (SHEDS-ATOX; Isakov et al., 2009), and wood preservatives (SHEDS-Wood; Zartarian et al., 2005; Zartarian et al., 2006; Xue et al., 2006; FIFRA SAP, 2004). This manual, however, focuses only on the residential module version 4.0 of the SHEDS-Multimedia model (hereafter SHEDS-Residential).

For background information, the SHEDS-Multimedia model development history, general overview of the SHEDS-Residential technical methodology, and summary of features and advantages of SHEDS-Residential version 4.0 are given in the following sections. After the Introduction, the manual describes details of the model structure and technical approach, model inputs, model outputs, sensitivity and uncertainty analysis capabilities, quality assurance, references. Appendices are provided for SHEDS terminology, code modules and their functions, description of input files, and a code verification simulation.

1.2 SHEDS-Multimedia Model Development History

SHEDS-Multimedia related research has been in development since 1998. Several publications are available (Xue et al., 2010c; Xue et al., 2006; Zartarian et al., 2006; Hore et al., 2006; Zartarian et al., 2000) and numerous technical presentations have been made at various national and international conferences and workshops. In 1999, ORD/NERL exposure modeling researchers developed the SHEDS-Multimedia (formerly called SHEDS-Pesticides) version 1 methodology and SAS code. This version of the model included dermal and non-dietary ingestion exposure routes; a micro-activity approach (i.e., estimating each sequential exposure for skin-to-surface, hand-to-mouth, and object-to-mouth contacts); lawn, garden, indoor crack and crevice, and indoor broadcast scenarios; 1-stage Monte Carlo sampling for estimating variability in population exposures; and a simple 3-compartment pharmacokinetic model. Version 1 was applied to a chlorpyrifos case study and a paper was written and published in 2000 (Zartarian et al., 2000).

Development of SHEDS-Residential version 2 (then called SHEDS-Pesticides 2; coded in both SAS and C++) began in 2001. This extended version 1 to a full aggregate model including the inhalation and dietary ingestion routes; switched to the macro-activity (transfer coefficient) approach for dermal and non-dietary ingestion exposure; included 2-stage Monte Carlo for assessing uncertainty as well as variability; and included research from the NERL University Partners (model evaluation from EOHSI's 10-home field study and pesticide background levels from LBNL's fate and transport model). SHEDS-Pesticides version 2 was applied for an October 2001 Aggregate Residential Exposure Model Comparison Workshop organized by EPA's Office of Research and Development and Office of Pesticide Programs. This workshop involved collaboration with other aggregate exposure modelers (CARES, Calendex, Lifeline) to assemble common inputs and scenarios, and compare model results. In 2002, SHEDS-Pesticides version 2 was reviewed as part of the ORD UPA (University Partnership Agreement) Peer Review. NERL began preparing for OPP's FIFRA SAP review of SHEDS-Pesticides version 2, then diverted resources to assist OPP's Antimicrobials Division develop a scenario-specific version of SHEDS for estimating exposure and dose of children to wood preservatives (SHEDS-Wood), for an application to CCA-treated playsets and decks.

From 2002-2004, SHEDS research mostly focused on SHEDS-Wood development and application, and methodologies that would be included in SHEDS-Residential (Xue et al., 2004; Xue et al., 2006; Zartarian et al., 2005; Zartarian et al., 2006). In 2004 when the CCA exposure assessment was being finalized, the extension of SHEDS-Residential version 2 to a more modularized version 3 began. In 2005, the focus of the SHEDS developers returned to SHEDS-Residential version 3 when NERL and its contractor Alion developed a SHEDS-Multimedia (aggregate residential) version 3 GUI, User Guide, and Technical Manual. In 2006, SHEDS-Pesticides was renamed to SHEDS-Multimedia since it is designed to accommodate other multimedia, multipathway chemicals (for both residential and dietary modules). The SAS code, GUI, and documentation were completed in the summer of 2006 for testing by several groups within and outside of EPA. The SHEDS-Multimedia version 3 (aggregate residential module) materials were further tested and externally peer reviewed by OPP's FIFRA SAP in August 2007. At that meeting, peer consult was obtained on the SHEDS-Dietary module and other plans for the cumulative residential version 4.0 algorithms.

Since the August 2007 FIFRA SAP meeting, SHEDS-Multimedia version 4 was developed, considering discussions and comments from that SAP (FIFRA SAP, 2007). SHEDS-Multimedia version 4 consists currently of the SHEDS-Residential module version 4.0 and the SHEDS-Dietary module version 1.0. In the next iteration of SHEDS-Multimedia, these two modules, which are currently standalone modules that have been "soft linked" for a permethrin pesticide case study, will be merged together, using a peer reviewed linkage methodology. This Technical Manual updates the SHEDS-Residential version 3 Technical Manual to reflect the model changes from version 3 to version 4.

1.3 SHEDS-Residential Version 4.0 Overview

This section provides a general description of the SHEDS-Residential version 4.0 model organization, structure, inputs, and outputs. Greater detail is provided on these topics in subsequent sections.

The primary function of the SHEDS-Residential version 4.0 model ("SHEDS" hereafter in this manual) is to estimate the exposure of a population to one or more specified chemicals after they have been inhaled, ingested (by mouthing of hands or objects), or dermally contacted in a residential setting. SHEDS uses the Monte Carlo statistical method to simulate a population of individuals based on the U.S. Census. These individuals are not specific persons, but are stochastically created synthetic persons whose collective properties reflect the simulated population and input distributions for exposure-related variables. For each individual, SHEDS constructs a sequence of activities, media concentrations, and the resulting exposures over the selected simulation period, which may range from one day to a year or more (although simulation time steps can range from 1 minute to 1 hour within a day). These individual exposure time series may be stored or exported, or aggregated over time to give time-integrated or time-averaged exposures (Figure 1-1). They may also be input to a dose model, either internal or external to SHEDS, to follow the fate of the chemical after it enters the human body. Exposure is defined in this model as the contact between a chemical agent and a simulated human target at the external body surface, either the skin surface or the oral/nasal boundary. Dose is defined in this model as the amount of chemical that enters the target after crossing the exposure surfaces.

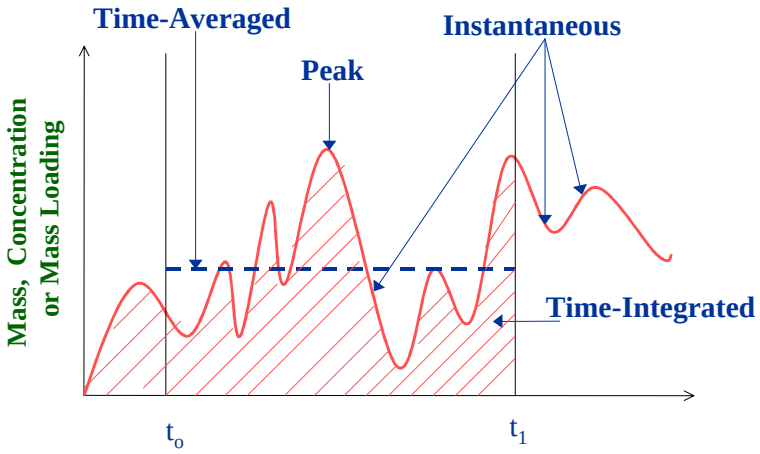


Figure 1-1. Hypothetical exposure profile for an individual.

SHEDS can be run as a one-stage Monte Carlo model or as a two-stage Monte Carlo model, allowing the estimation of variability, uncertainty, or both (see Figure 1-2). A standard SHEDS model run, called a "variability run," involves generating exposures for a random sample of individuals from the target population using Monte Carlo sampling (Cullen and Frey, 1999), thereby determining the exposure distribution for the population. Due to the stochastic nature of the model, no SHEDS individual will represent any one specific real person. Therefore, while pairing SHEDS exposure estimates with personal exposure measurements at the level of specific individuals has been done (Hore et al., 2006), it is not generally advisable.

SHEDS-Residential is coded in SAS version 9.1; the user must have this version (or later) of SAS installed prior to the installation of SHEDS (see details in the User Guide, Isaacs et al., 2010a). On a typical personal computer, a SHEDS variability run of 1000 persons will take about one hour and produces between 200 megabytes and 5 gigabytes of output, depending on the choices for saving output. These numbers should scale linearly in the number of persons, until limited by computer resources.

In a two-stage uncertainty run, a series of variability runs are performed, with the input distributions modified between each variability run to represent uncertainty in the input parameters of the

variability runs. This allows for the construction of confidence intervals for various points on the variability distribution (see Uncertainty Analyses section for details). SHEDS may also be run in a special mode designed for sensitivity analysis, which estimates the influences that various input variables have on exposure and dose statistics (see Sensitivity Analyses section for more details). Sensitivity and uncertainty runs are more computer resource intensive than variability runs, and large runs may take several days to complete.

SHEDS-Residential version 4 uses an overall random number seed to control the generation of random numbers. Repeating an earlier run with the same seed leads to identical output. Furthermore, limited changes can be made between runs (while keeping the seed the same), which leads to certain shared features among the runs. This is the basis for certain types of sensitivity analysis.

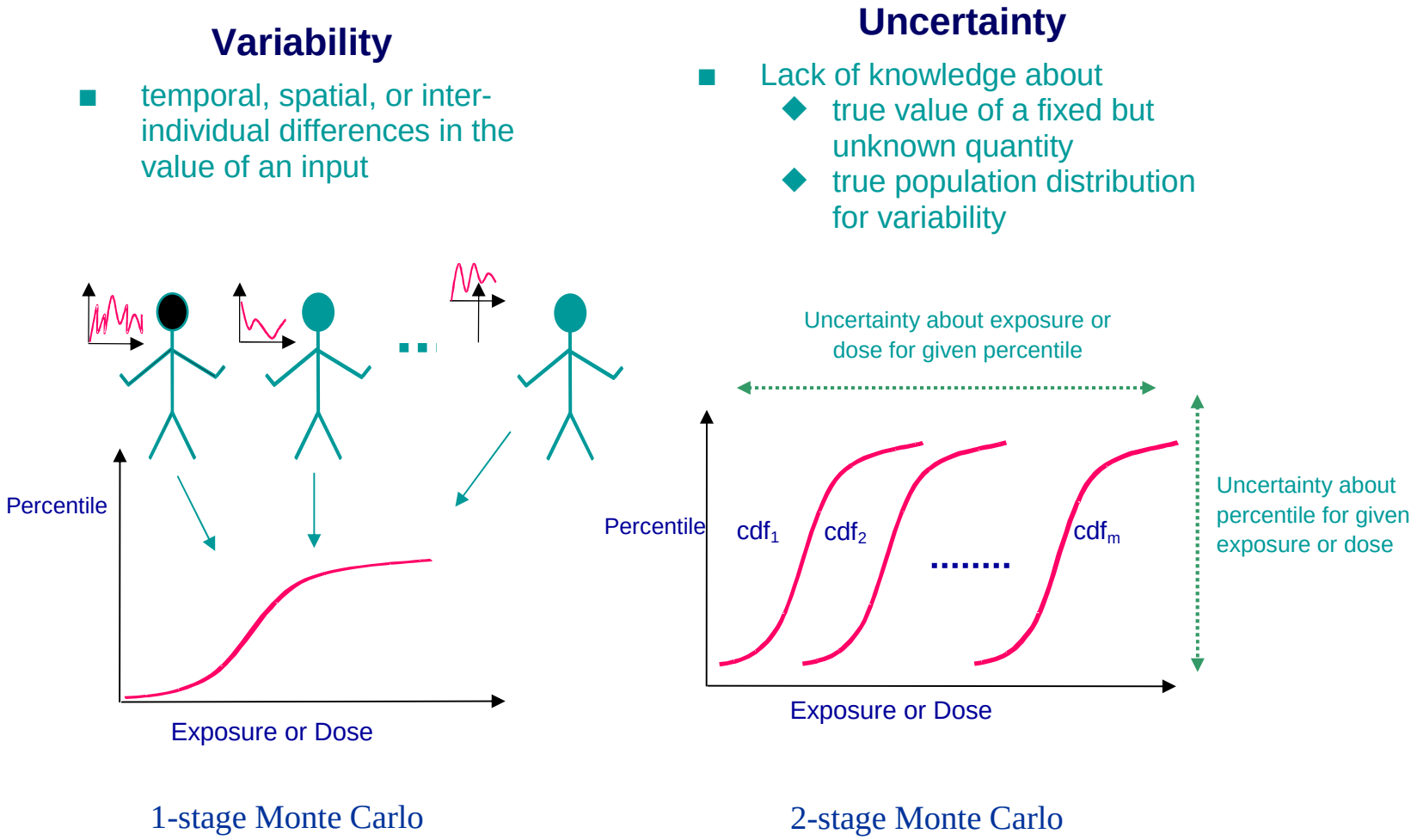


Figure 1-2. Monte Carlo sampling in SHEDS.

The fundamental modeling unit in SHEDS is the individual. Since each individual is generated as a representative random sample, the rules for determining the characteristics and the exposure for each individual are simply repeated for as many persons as the user wishes to model. One individual differs from another only because there is stochastic variation when randomly sampling from distributions. For each individual in a SHEDS-Residential run, the following general steps are applied:

1. Randomly select the age, gender, and other demographic properties of interest, given the distribution of the target population.
2. Generate a longitudinal activity diary, which indicates the sequence and duration of activities and locations for that person. For the residential module, these are based on sequential time-location-activity diaries from EPA's CHAD database.
3. Generate concentration time series for each potential contact medium (e.g., indoor air, indoor smooth surfaces, indoor textured surfaces, outdoor lawn, outdoor vegetable garden, and pets). These concentrations may depend on the usage patterns for household pesticides or other chemicals.
4. Simulate the contacts between the individual and the affected media. These depend on the diary activity/location information and contact probabilities derived from user-specified inputs.
5. Calculate pathway-specific exposure time series for the individual, using the results of the prior two steps and user-specified distributions for exposure factors.
6. Generate an approximation for the components of the intake or absorbed dose time series and export these for use in a PBPK model.
7. Time-aggregate to daily totals and use the simple built-in pharmacokinetic model in SHEDS to obtain estimates of blood and urine levels.
8. Determine margins of exposure (MOEs) from the exposure and dose results.
9. Output summary statistics for this individual.

To obtain population estimates SHEDS repeats this process for an individual many times using Monte Carlo simulation. Once the model run is complete, the user may view summary tables and graphs of the results. SHEDS includes four exposure pathways (dermal, hand mouthing, object mouthing, and inhalation). Each of these is reported on each diary event, with typically 12,000-20,000 diary events per year, per individual. A **diary event** is a particular activity performed by an individual in a given location; events are the base time increments of the CHAD database diaries that are used in SHEDS; they range in length from 1 minute to 1 hour. When output from just one individual is examined, it is usually plotted as a time series. When collectively examining output from a population, usually each person is summarized by a single value (often some average exposure or dose over the simulation period), and the population distribution of this statistic is generated. An overview of the SHEDS-Residential version 4.0 methodology is given in Figure 1-3 .

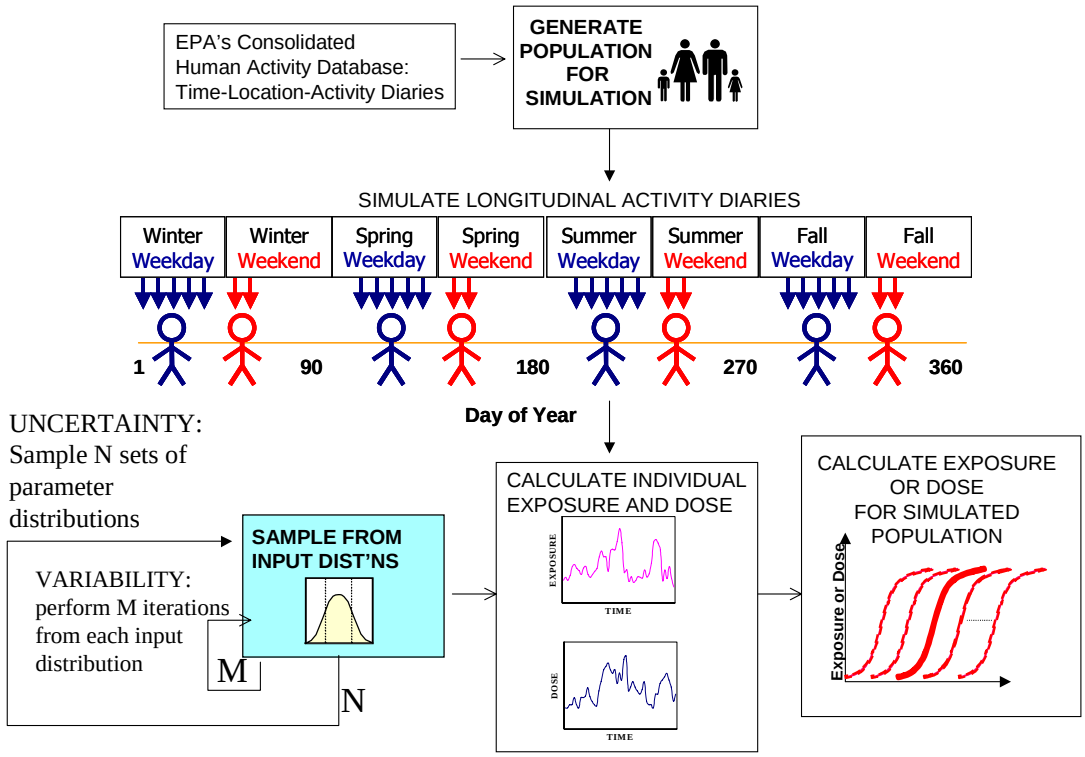


Figure 1-3. Overview of SHEDS residential methodology.

1.4 Features and Advantages of SHEDS-Residential Version 4.0

Key features of the cumulative version of SHEDS-Residential version 4.0 include the following:

- physically-based, probabilistic exposure model coded in SAS 9.1 with modularized code, flexible user input files, user-friendly GUI;
- generates variability and/or uncertainty distributions of exposure and/or dose;
- simulates route-specific (hand-to-mouth, object-to-mouth, dermal, inhalation) exposures/doses, either aggregate (single chemical) or cumulative (multiple chemicals);
- models both handler (during application) and post-application exposures;
- time series approach uses EPA's CHAD macroactivities as basis for longitudinal activity patterns and time steps; accounts for variability within a day;
- includes co-occurrence of chemical usage;
- chemical application scenarios include residential indoor crack and crevice, pets, indoor fogger, lawn, and garden (for different application methods);
- 3 options for source-to-concentration module: read externally prepared time series; sample from distributions based on time since application; use built-in decay/dispersion model;
- 2 options for dose modeling: export exposure time series to PBPK model; built-in simple PK model;
- 2 options for specifying surface concentrations: as mass per unit area ($\mu g/cm^2$), or mass per unit mass ($\mu g/g$) which is intended for chemical measurements in dust or soil samples;
- 2 options for dermal exposure: transfer coefficient or transfer efficiency;
- code permits correlation of randomly sampled model inputs;
- built-in decay/dispersion model includes treated and untreated room; applies to concentrations for smooth or textured indoor surfaces, air, lawn, vegetable garden, and pets;
- separates dermal hand and body exposure profiles and links hand-to-mouth ingestion with dermal exposure;
- accounts for sequential activities, dermal replenishment, and removal processes (hand-to-mouth, bathing, hand washing, absorption);
- calculates margins of exposure (MOEs); and
- has 3 built-in methods of sensitivity analysis.

Potential future enhancements/features of SHEDS-Residential could include the following: background levels indoors (due to track-in or air infiltration); modeling exposure away from home; refined linkage with other models (e.g. fugacity); testing of longitudinal analyses; model evaluation; and linkage of model simulations with geographic information systems (e.g., for community level exposure assessments).

1.5 Cumulative (multi-chemical) modeling in SHEDS

An important feature of SHEDS version 4 is the ability to model exposures of the same individuals to multiple chemicals, whether these chemicals are present in the same application scenarios

or not. A detailed explanation of the SHEDS multi-chemical methodology is presented later in the Methods chapter. In this introduction, a general overview is provided.

The SHEDS version 3 code consisted of two major nested loops: the inner one simulated one person per pass, while the outer loop varied the input data for sensitivity or uncertainty analysis. Version 4 maintains this structure, but adds a third loop over chemicals. Logically, the chemical loop would be the innermost loop, since the same simulated persons are being exposed to each chemical. However, the version 4 code currently places the loop over chemicals as the outermost loop. This structure allowed the greatest overlap with the pieces retained from the version 3 code. Another advantage to this structure is that the number of internal coding variables does not increase with the number of chemicals. Otherwise, SHEDS would more rapidly reach SAS limits or the hardware limits of the host computer.

Another new feature in version 4 is the explicit control of random number seeds. By using the same seeds for personal demographic data in each pass through the chemical loop, the same individuals are generated on each pass, with the same activity diaries, scenario usage, and contacts with media. On each pass through the chemical loop, the chemical is identified and input data specific to that chemical are extracted from the input files. From that point forward, until the end of this pass through the chemical loop, SHEDS operates as in version 3, in which all the internal variables are geared toward a single chemical. For example, the variable that records new dermal exposure does not need a chemical-specific suffix to distinguish it from the new dermal loading for another chemical. The same generic variable is used for each chemical in turn.

At the end of each pass through the chemical loop, the exposures and doses are appended to the output files along with a variable containing the chemical name. For example, if a model run consists of 365 days, 1000 persons, and 10 chemicals, the exposure output files would contain 3.65 million records, one for each combination of day, person, and chemical. The first 365,000 records would be for the first chemical, and so on. When graphing, summarizing, or tabulating the output one chemical at a time, it is straightforward to select the portion of this output file corresponding to the selected chemical, and then the same code and analysis tools may be used as for a single chemical (aggregate) model run.

The most important feature of the cumulative modeling in SHEDS is that when two or more chemicals of interest occur in the same application scenario, there is a very strong relationship in the chemical concentration time series and in the resulting human exposure. The peaks in the concentration time series for such chemicals will occur on the scenario usage dates (application dates). Extensive contact with the affected media will result in high exposure to all the chemicals present, whereas little contact will result in low exposures to all chemicals. Of course, this does not apply to chemicals found in different locations; for example, if chemical X is used indoors while chemical Y is used on the lawn, it is possible to have high exposure to one but not the other.

SHEDS version 4 assumes that the chemicals are non-interacting, either in the environment or once they are on/inside the person. The code loops over chemicals; the same person engages in the same activities with the same sequence of media contacts in each pass, except that one chemical has been replaced by another. Since all knowledge of the amount of the other chemicals is lost, there is no possibility of interactions between chemicals in this model structure.

1.6 Changes from SHEDS-Residential version 3 to version 4

The following lists provide a summary of the main changes from SHEDS-Residential version 3 to version 4. Most of these changes are based on review comments from the 2007 SAP review (FIFRA SAP, 2007); some are based on considerations that arose during review of the model, and the desire to simplify the code and reduce the number of user-specified inputs. Note that for the case study described in

Section 7.3, these collective differences did not yield substantial differences between version 3 and version 4.

Capabilities added in version 4

- **Multichemical runs**: Version 3 could be adapted to run different chemicals, but since separate model runs were required, the set of simulated individuals was different for each chemical. Version 4 runs each chemical with the same persons, activity diaries, and scenario usage patterns, producing correlation in exposure across chemicals.
- **Random number reproducibility**: This allows previous runs to be repeated exactly. It also allows slight modifications of earlier runs while preserving other details, which is useful in sensitivity analysis.
- **Handler exposures**: Dermal and inhalation exposures are now made for non-professional handlers (chemical applicators).
- **Scenario-specific treated areas:** The size of the treated area for indoor applications can be varied by scenario. Some may be applied to the entire house while others are applied to a small area. This directly affects contact probabilities.
- **MOE calculations**: Margins of exposure (MOE) are now calculated and reported.
- **New longitudinal diary assembly method**: The new method based on the D and A statistics is now available, along with the 8-diary method from version 3.
- **Event-level output for dose modeling:** Event-level time series of the amount of chemical crossing the exposure-dose boundary may be saved for all persons, which can be used as input to a PBPK dose model.
- **Bi-phasic decay of air concentration**: For many indoor pesticides, there may be a high initial air concentration that drops by a factor of 100 or more during the first day. After this, decay is gradual and comparable to the decay in other media. SHEDS now requests both initial air concentration and the value after 24 hours as user inputs.
- **Empirical data:** The user may enter empirical data for any randomly-sampled input requiring a user-specified distribution.
- **New options for uncertainty inputs**: All user-specified inputs (except for job settings) may be given uncertainty distributions, and additional distributional properties may be modified beyond the options available in version 3.
- **New distribution truncation options**: Randomly sampled values that fall outside the specified bounds may now be effectively resampled, or alternatively allowed to 'pile up' at the bounding points as in version 3. The new option avoids having extra values at the bounding points. Also, the bounding points may be specified either as values or as percentiles of the distribution.
- **Sobol's method of sensitivity analysis**: This technique has been added to the existing forms of sensitivity analysis retained from version 3.
- **Running selected persons from earlier run**: When reproducing another run, the user may request output only for a selected list of persons, rather than everyone. This is faster and allows for more detail to be output on each selected person.

- **Saving all intermediate variables for a selected person**: In version 3 this only existed for the final person simulated in a run, since (to save space) many of the data sets were re-used for the next person. In version 4, intermediate variables can be obtained for any given person by requesting a previous run to be repeated with a single selected person (as in the previous bullet).
- **Sensitivity analyses can be conducted through the user interface**. Percentile scaling, Pearson and Spearman correlation, and Sobol methods are all options.
- **Uncertainty runs through the user interface**. Uncertainty runs require uncertainty clouds to be supplied by the user. It is suggested (but not required) that these be produced using the bootstrap method described in Chapter 7.

Algorithms modified for version 4

- **Exposure – dose boundaries**: The exposure surfaces are now at the external surface of the human body (i.e. outermost skin layer and oral/nasal boundary). "Dose" measures the amount of chemical crossing this boundary. This change required deletion of certain version 3 algorithmic details that applied to internal body chemistry, and addition of a dermal binding variable related to transfer from the external skin contact boundary into the *stratum corneum*.
- **Chemical phases**: Chemical concentrations may still be specified separately in dust or soil, but are now combined with surface loadings in the environment before contact occurs. The same transfer equation (either transfer coefficients or transfer efficiency) is used for all forms of the chemical.
- **Human-chemical contact logic:** Fewer input variables are required than in version 3; see section 1.6.1 below. Multiple surface media may now be contacted on each diary event. This may be called "event splitting" since each contact duration may be shorter than the event duration. See section 2.6 for details.
- **Blood and urine doses:** These are now calculated on a daily basis rather than on each event. For greater detail in modeling the chemical inside the body, the SHEDS output files can be used as inputs to a PBPK model.
- **Maximum dermal loading:** This is now a fixed (unchanging) limit for each person.
- **Hand washing:** This may occur only once in each clock hour.
- **Transfer rates:** The rates of chemical transfer onto the skin are now reduced when pre-existing dermal loading is present.
- **Mets adjustments:** An additional adjustment is made to each Mets distribution (metabolic ratio of energy expenditure for an activity to the resting rate) which relates its mean value with the current person's maximum allowed mets value.
- **Correlation of inputs:** Spearman correlation has been extended to a greater variety of inputs. Now only inputs related to usage are excluded, but those are still subject to co-occurrence logic (which can effectively produce correlation in usage patterns).
- **Determination of concentration in indoor untreated air:** Instead of following the pattern of the carpet and hard floor (in which there is a long delay before the concentration reaches its maximum in the untreated area), the untreated air concentration

will quickly reach a dynamic equilibrium between the treated air and the outdoors (which has zero concentration) due to the time scale of the air exchange rates. This effect was seen in the results from the fugacity model developed for use with SHEDS.

- **Expanded CHAD activity diary database:** Several studies have been added to the CHAD database, making it ~50% larger than the one used in version 3. Also, the default set of rules for accepting diaries for use in SHEDS has been slightly modified.

Options removed in version 4

- **Direct soil ingestion**: Version 4 allows only the indirect option which uses hand loading and hand-to-mouth transfer. This is necessary since soil is no longer distinguished from other components of the hand loading. See 1.6.2 below for more details on the reason for this change.
- **Event-level resampling of random variables:** In version 4 the most frequent resampling rate is hourly. This simplifies the code, especially in connection with random number reproducibility.
- **Indoor background concentration from outdoor sources**: Version 4 calculates chemical present indoors that resulted from earlier indoor usage in each house, based on the usage patterns assigned to that house. It does not allow for chemical to enter by track-in or air infiltration, which to be logically implemented would require a connection between indoor and outdoor concentrations.
- **Shortening of activity diary events**: Version 3 allowed the user to reduce the maximum event duration below 60 minutes. Version 4 replaces this option by allowing multiple contact media within each event (called "event splitting"; see section 2.6).

In addition to the new features, the version 4 code is more readable than the version 3 code. The variable naming has been standardized, and less macro language code is used.

1.6.1 Input file for contact probability variables

Both version 3 and version 4 of SHEDS-Residential simulate human contact with the environmental media containing the chemical agent. Version 3 had a list of 18 input variables for this purpose, most of which were conditional probabilities such as "probability of being in a treated area while at home and awake". There were two difficulties with this: the number of such inputs was rather large, and some users found the variables difficult to understand or estimate.

One of the extensions made in version 4 was to allow each indoor scenario to have its own treated area. Some scenarios may treat the entire house, but others will usually be applied to a few specific locations. When multiple scenarios are modeled in the same run, variables such as "probability of being in a treated area while at home and awake" would have to become scenario-specific, which would increase the number of inputs beyond the 18 required in version 3. Hence, a method for simplifying the specification of contact probabilities was developed for version 4.

One change in version 4 was to reduce the number of areas in the house from three to two. In version 3 there was a treated area where the chemical was directly applied, an untreated area to which the chemical could disperse, and a third area with zero chemical concentration (too far from the treated area to receive appreciable dispersion). In version 4 this third area is eliminated. The user can still create

areas with zero concentration in the modeled house by assigning zero dispersion into the untreated area. Otherwise, the untreated area in version 4 should be considered to be the average of all the non-treated parts of the house. In a real house, concentration would vary along a gradient, as areas adjacent to a treated area will receive more chemical than those farther away. So in both version 3 and version 4, the concentration of the "untreated area" represents the average of several areas. In version 4, this average includes the areas with zero concentration.

Another change in version 4 is that the chemical in dust or soil is no longer distinguished from chemical in other forms, in terms of the potential for contact. In version 3, on a single indoor diary event it was possible to contact dust only, residue only, or both, or neither. In version 4 the chemical in dust is effectively mixed with the chemical in other forms to produce an overall chemical loading on the floor. Then if contact occurs, part of this total loading is transferred to the person. However, this requires fewer contact-related input variables since the dust and residue contacts do not have to be determined separately.

A third change in version 4 is that the remaining input variables relating to contact are expressed in forms that are more easily understood. For example, the variable F_carpet is simply described as "the fraction of the house that is carpeted", but a better description would be "the fraction of in-home time spent in carpeted areas, while awake". Version 4 uses F_carpet to internally determine the probability of contact with a carpeted floor.

With all the above changes, SHEDS version 4 has replaced the 18 former inputs with just one (f_AreaTreated) for each indoor scenario, plus four inputs (f_carpet, f_lawn, f_garden, and f_pet) that are not specific to any scenario.

1.6.2 Dust and soil ingestion in version 4

SHEDS-Residential version 3 allowed the user to choose either direct or indirect estimation of dust and soil ingestion. The direct option has been discontinued in version 4. Chemical concentrations in dust and soil may still be entered in version 4, but these are converted to the equivalent surface loadings before being contacted. The two reasons for eliminating the direct soil ingestion are model simplification and model consistency.

In terms of model simplification, SHEDS version 3 required both more input variables and more internal model variables. The model had to calculate two concentration time series in each potential contact medium, one for chemical in dust/soil and the other for the remaining chemical (assumed to be in "residue" form). Separate transfer equations were used for the soil/dust and residue to the skin, and separate dermal absorption rates were allowed. The basis for these calculations was the SHEDS assumption that dermal exposure and dose are determined by both the human macroactivity pattern (basically, where one spends time) and the microactivity pattern (that is, the amount of "touching" of contaminated media). However, the direct dust/soil ingestion algorithm is not really compatible with this approach, because the amounts are determined by sampling from the input distribution for daily totals and are not directly connected with the microactivity pattern.

The consistency issue arises from the fact that version 3 had both residue and dust/soil loadings on the hands and modeled the hand-to-mouth transfer of the former in a mechanistic manner, counting the number of hand-mouth contacts, the area affected, and the transfer rate. However, when direct soil ingestion was used, this approach was not used for the dust/soil portion of the hand loading, to avoid double counting of the ingested amount. That is, the dust/soil ingestion can be determined with or without first calculating the amount getting on the hands. Version 3 calculated the ground/floor-to-hand

transfer of dust/soil, but did not continue with the hand-to-mouth transfer, which was problematic. One might conclude, for example, that hand washing or bathing could be an effective reduction strategy for limiting ingestion of chemical residues, but would have no effect on ingestion of chemical in dust or soil. Also, since the chemical on the hands is a mixture of both forms (residue and dust/soil), it would be logical that the ingested amounts from these two forms should be strongly correlated. This connection was absent in version 3. Another non-dietary ingestion pathway in version 3 was object mouthing – for example, toys that acquired chemical from being left on the floor. But since the chemical on the toys was not distinguished between residue and dust/soil form, there was no restriction on object-to-mouth transfer. Thus, version 3 double-counted the dust-soil ingestion via this pathway.

Ozkaynak et al. (2010) demonstrate that the indirect dust and soil ingestion algorithms in SHEDS give comparable results (when averaged over time) to the direct approach. In SHEDS version 4, only this indirect approach is supported. That is, all non-dietary ingestion occurs as the result of two processes: hand-to-mouth transfer or object-to-mouth transfer. The amount of chemical that gets on either the hands or the objects represents the combined amount of chemical in all forms, including chemical in dust or soil.

2 MODEL STRUCTURE AND APPROACH

This section covers the SHEDS-Residential version 4.0 modeling methodology in detail. SHEDS inputs and outputs are discussed in detail in subsequent sections. The fundamental structure of SHEDS is the stochastic simulation of activity and corresponding exposure for one individual at a time. Collectively, a large number of such individuals represent the target population. The population of individuals is simulated using Monte Carlo sampling. Simulation of individuals is described first, followed by population simulation. Figure 2-1 provides a general overview of input sources, model structure, outputs, and potential applications of SHEDS results. Note that Figure 2-1 is not intended as a detailed explanation of the operation of the underlying SAS code.

SHEDS-Residential may be run either via the user interface (GUI) or in batch mode. For batch mode runs, all of the inputs must be pre-defined by editing the appropriate SAS data sets. The user starts the batch run from the Run dialog on the Windows Start menu. The job then runs in the background and the usual SAS windows do not appear. Batch mode is required for all uncertainty and sensitivity runs, and is useful for other jobs that require a long run time. Detailed instructions on setting up batch runs are in the SHEDS User Guide.

SHEDS version 4 was designed to be very flexible; thus, the amount of "hard-wiring" was kept to a minimum. The data used in the model fall into three categories: user input, default input, and values embedded in the code. The variables in the first category are discussed in detail in the section on model inputs. Default inputs include the CHAD database, the population statistics from the U.S. Census, the height and weight tables based on NHANES III, the list of application scenarios and contact media, and standard age groupings. Without changing the model code, these defaults could be altered if the user supplied alternative data sets, but this feature is not covered in these manuals. Values embedded in the code are used in the following: specification of the seasons; mapping of CHAD codes to SHEDS location categories; basal ventilation rate equation; body surface and hand surface area equations; maximum METS, maximum oxygen debt, and recovery time formulae; the post-application transition days; the unified METS adjustments; and the rules for construction of longitudinal diaries from eight one-day diaries.

2.1 Defining Exposure Scenarios and Simulation Information

The following section briefly summarizes the steps needed to define a model run. More details on the algorithms and internal operations of the model are provided after this section. See the SHEDS-Residential User Guide (Isaacs et al., 2010) for explicit detail on using the SHEDS-Residential graphical user interface (GUI) for entering data.

The SHEDS user first selects the overall constraints on the model run. The first is the mode of operation: variability, sensitivity, or uncertainty run, and the number of individuals for the simulation specified. Note that sensitivity and uncertainty runs cannot be run through the GUI; they can be run with the SAS code in batch mode.

The simulation period must be defined, using a starting date and the number of days in the simulation period. SHEDS follows the real calendar, so a one-year simulation starting January 1 could start on any day of the week and could contain either 365 or 366 days, depending on which year is selected. The simulation period is any whole number of days. Simulations over one year in length can be run, but note that the individuals in this version of SHEDS are assumed to remain the same age throughout (in part due to lack of available exposure-relevant information on aging an individual). If the

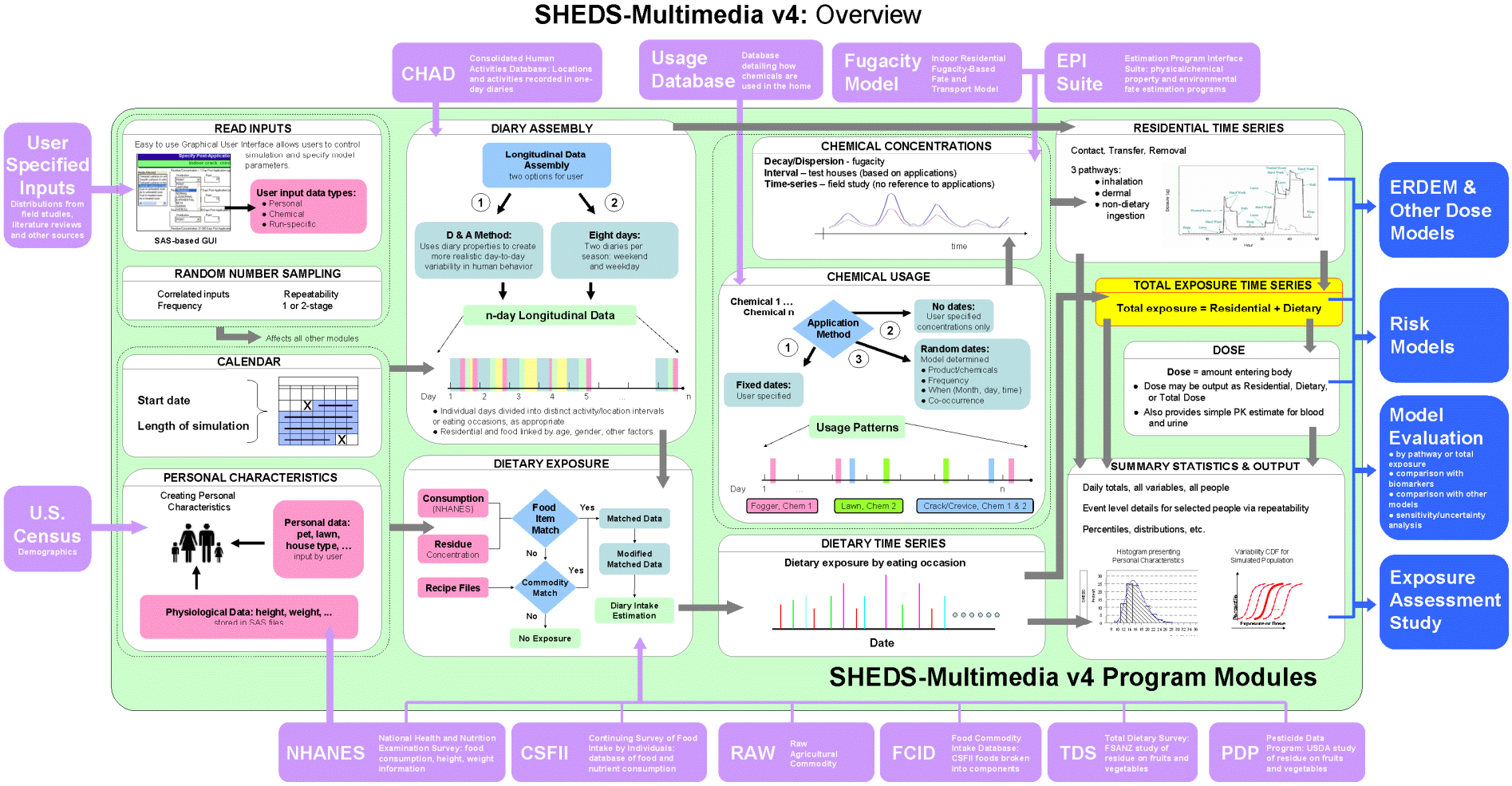


Figure 2-1. General overview of SHEDS-Residential exposure model

concentration time series option is used, then the simulation period may be shortened to match the selected concentration data.

Next, the target population is defined in terms of its age and gender. The user selects the allowed age ranges separately for males and females. The number of individuals to be run is then selected. Since these individuals are randomly generated and are representative samples of the target population, it is possible to combine outputs from several model runs that share the same input settings. A larger number of individuals (i.e., population size) results in a smaller stochastic variation in the exposure distribution. As a rough guide, exposure statistics are generally reproducible within plus or minus 10% when the sample size is around 1000 simulated persons, or plus or minus 1% for 100,000 simulated persons. While computer resource limitations may make a single run of 100,000 persons impractical, if this number were desired one could perform 10 runs of 10,000 persons each (making sure that each run uses a different random number seed, but otherwise using identical inputs). However, if this is desired, the combining and joint analysis from the various runs must be performed outside the SHEDS interface. In the standard mode of operation supported by the interface, the entire set of individuals to be analyzed must be generated in a single model run.

There are several other choices that must be specified to define the model run. In particular, there are multiple ways to determine the chemical concentrations in the various contact media. One of these options is for the user to supply concentration time series directly in an input file (e.g., data from a measurement field study or output from a source-to-concentration model). The other options first require the determination of chemical usage. This is based on a "scenario orientation" in which the user defines any number of distinct application scenarios, which are combinations of chemical formulation, method of application, and target location. There are two methods for determining the dates of chemical usage: fixed dates ("User-Specified") or variable dates ("Model-Determined"). The user may specify a set of dates on which the chemical application scenarios are used (the 'fixed dates' mode), in which case all simulated persons share these usage patterns. Alternatively, the model may randomly generate usage patterns using user-supplied probabilities (the "Model-Determined" mode), in which case each simulated person is likely to be different. In this mode, some persons might have no chemical usage at all. With either method of defining usage patterns, there are two ways to determine chemical concentrations at subsequent times. The decay/dispersion option allows the chemical concentration to decay over time in treated areas, while the chemical may slowly disperse into untreated areas. The intervals method selects concentrations from user-specified distributions that depend on the amount of time that has passed since the chemical was last applied. These methods are described in more detail below. The final task for the user is to set the distributions for the input variables.

The following sections describe the modeling steps for creating an individual, assigning diaries to them, determining the chemical concentration time series (which may depend on usage patterns), generating the human-chemical contacts, and calculating the resulting exposure (and optionally, dose) as functions of time. These steps are then repeated for each individual using Monte Carlo sampling to simulate a population from which output information can be obtained.

2.2 Generating randomness in SHEDS

SHEDS is a stochastic model that relies on the random sampling of distributions to construct simulated individuals and their exposure sequences. Each such individual is a representative sample of the target population, as is a collection of such individuals (if there are no duplicates). SHEDS models one individual at a time before proceeding to the next, and there is no relationship between different individuals in the same run. Effectively, the code “forgets” everything about the previous individual when another is started, and a completely new set of random numbers is drawn for each person.

Within an individual, a set of exposure time series are generated to cover the simulation period, typically one year. While some properties of the individual and their environment vary randomly over time, many are either constant or deterministically follow from prior values. Therefore, an individual's daily exposures are not independent of their exposures on earlier days.

A major new feature in SHEDS version 4 is the option of reproducing random number streams from prior runs. Furthermore, the same random numbers may be used even if the model inputs have been changed in certain ways. This allows new options for sensitivity analysis.

Random numbers in SAS (like in nearly all computer languages) are really pseudo-random, and actually follow deterministically from a "seed" value. If the same program is run again with the same input settings and seed value, then the same set of random numbers will be produced. The Rand function in SAS has its seed initialized or set just once in each SAS data step. In SHEDS version 3 this initialization was performed by calling the computer's clock and using the returned time. Every SHEDS v.3 run therefore used different seeds and produced different results, even if all the job settings were the same.

SHEDS version 4 starts with one user-supplied seed. If the user leaves it at zero, then the clock supplies this seed. Either way, the seed is then recorded on the job settings file so the run may be repeated. From this, the model generates a set of three new seeds for each individual in the run, which are recorded for later use. Three seeds are generated per person because the code has three data steps per person that generate random numbers, and each one requires a different seed.

The storage of personal seeds allows for new functionality in version 4. For example, it is possible to request selected individuals from a previous run. Suppose a large run of many thousands of persons was made, and a few showed unusual results. SHEDS normally overwrites most of the details of the exposure calculations when proceeding to the next person, for space considerations. But in version 4 one can request that the model retain the full output from selected individuals from a previous run. This is accomplished efficiently by using the personal seeds to reproduce only the selected individuals, while not modeling the skipped individuals at all. If the input settings have not changed and the same three personal seeds are used again, then exactly the same person with the same exposure time series will be produced each time the model is run.

All the random numbers produced in SHEDS start out as samples from a uniform (0,1) distribution. These are called u_variables, where the prefix "u_" indicates the uniform distribution. For example, a uniform random variable called u_age is later converted to an actual age using the population data. For nearly all random variables, the conversion is rank-preserving. Thus, if u_age=0.37, then the person is assigned the age at the 37^{th} percentile of the population age distribution.

The three data steps that generate random numbers are the following: the data step u_base produces the random numbers for the age, gender, body weight, and diary selections. The data step u_simulation produces all other random numbers not associated with specific scenarios. These are either constants for the given person (such as height), or vary on predictable schedules (monthly, daily, or hourly). The data step u_products generates the remaining random numbers. Note that while some of the random numbers on u_products will be used later to select the usage schedule, but when they are generated these random numbers are not (yet) associated with particular simulation times. For this reason, SHEDS does not permit the correlation of random numbers on u_products. Any or all of the random numbers on u_base and u_simulation may be correlated with each other, at user request.

The u_base random variables may be reproduced in the SHEDS-dietary program, allowing matching individuals (by age, gender, weight, and diary variables) to be created in both models. The diary variables allow the longitudinal food and activity diaries to be correlated, creating a linkage between dietary and residential exposures.

2.3 Creating a Simulated Individual

2.3.1 Age Groups

Each individual modeled in SHEDS is assigned a specific year of age. SHEDS uses age groups for three distinct purposes, and the age groupings may differ for each purpose. First, the model samples activity diaries from pools spanning the diary age groupings. For example, if an 8-year-old girl is simulated, diaries from the pools for 6 to <11 year-olds are sampled to generate that child's longitudinal activity profile. Second, the model allows age-dependent model inputs; for example, hand-to-mouth frequency. Even though the set of activity diaries are the same for an 8 or a 9 year old child, other inputs could distinguish these ages. Third, the output or results from a model run may be grouped by age. This is a type of post-processing and is entirely at the user's discretion.

The residential module of SHEDS supports two sets of age groupings for diary pooling. The first set is: 1 to <2 yr, 2 to <3 yr, 3 to <6 yr, 6 to <11 yr, 11 to <16 yr, 16 to <21 yr, 21 to <30 yr, 30 to <50 yr, 50 to <70 yr, 70+ yr. The groupings for ages 1 to 21 years come from U.S. EPA (2005), which provides guidance on selecting age groups for monitoring and assessing childhood exposures to environmental contaminants. These are also being used in the updated EPA Child-Specific Exposure Factors Handbook (EPA, 2008).

The above list does not include children less than one year of age because of some special considerations related to this age group. First, though some are in CHAD, data availability in the form of time activity diaries is limited. The CHAD activity and location codes may need to be re-examined for this age group. The METS values and contact probabilities currently employed by the model might need to be changed, and in this regard, the question of data availability again arises. Another point is that children of this age change very rapidly, both physically and behaviorally, and the SHEDS model currently does not "age" simulated individuals.

The second set of age groupings for diary pooling is: 0 yr, 1 to <3 yr, 3 to <6 yr, 6 to <13 yr, 13 to <20 yr, adults 20 to <50 yr, females 13 to 49 yr, adults >=50 yr). These match the groupings currently used in the SHEDS-Dietary module. For consistency with the dietary module, age 0 is allowed in SHEDS-Residential. The reservations listed above are still relevant, but might be handled in part by making certain inputs conditional on age via the construction of "fine tuned" input data files. Some experienced SAS users may wish to attempt this, but this option is not currently available through the GUI.

For outputting results, the GUI allows the user to specify one age group at a time and then it calculates statistics for that group. The user may thereby re-group the simulated individuals in any manner.

2.3.2 Assigning ages to individuals

SHEDS includes nationwide population counts for each year of age and gender, from the 2000 census (U.S. Census Bureau, 2002a,b). Through the interface, the user may select any or all of the 20 age-gender cohorts to be in the target population. In batch mode the user may select any or all specific ages. SHEDS randomly assigns a particular age and gender to each individual, and these are recorded in the output files. Thus, it is straightforward to group the results into any desired age cohorts. For example, the collective exposure distribution for persons of ages 3, 4, or 5 years can be determined by selecting the simulated persons of those particular ages out of a larger run. The main concern is that the model run has to be large enough to ensure a sufficient number of people in each age cohort to allow for a good characterization of the exposure distribution. The issue of sample size becomes more acute for the younger cohorts, as they span fewer years of age and represent smaller fractions of the population.

Stratified sampling of age cohorts would mitigate this, but then it would be more difficult to construct a variability distribution for the entire population.

When a SHEDS individual is assigned a particular age and gender, then all inputs that depend on these variables will be chosen appropriately. The activity diaries will be selected from individuals belonging to the same age-gender cohort. It is not practical to match age exactly, as there would not be enough diaries to achieve this in all cases. Since some of the inputs are age-dependent, the exposure distribution for persons of a given age will differ from that of the general population.

2.3.3 Assigning other person-level variables

The 'OnePerson' module in the SHEDS code assigns age and gender to individuals, along with several other input variables that are sampled only once per person. These include personal characteristics such as height, body weight, skin surface area, basal metabolic rate, maximum metabolic rate, and characteristics of the person's home, such as whether or not they have a lawn, a garden, or a pet, and the areas of these objects. Some behavioral variables such as the mean number of hand washings per day are set once per person.

Each person is given a specific body weight sampled from a lognormal distribution, with parameters that are age and gender-specific. The NCHS/CDC Third National Health and Nutrition Examination Survey (NHANES III, 1988-1994) was the source of the data used to develop these distributions. At present the weights are not capped at minimum or maximum values. However, a sample of SHEDS-generated body weights for 10,000 adults gave a minimum of 35 kg or about 77 pounds, which is not unreasonable. The maximum was less than 300 pounds. Once body weight is assigned, other variables including height, skin surface area, and basal metabolic rate are sampled, using relationships that depend on age, gender, and body weight.

The distributions that determine most of the person-level variables are not user-adjustable, unless the user edits the SAS code directly. Three "persons" variables that do depend on user settings are whether each individual has a lawn, a garden, or a pet. For each of these, the user enters a binomial distribution, and a yes/no test is performed for each of these questions, for each individual. Other user-specifiable person-level variables include the areas of the house, lawn, garden, and pet; the fraction of indoor time spent in carpeted area; and the fraction of time outdoors (while at home) spent on the lawn and in the garden.

The *variables* input SAS data set contains information on every randomly sampled variable in the SHEDS model. It also specifies the sampling frequency (which may be once per person, once per year, month, once per day, or once per hour), the SHEDS module in which the variable is used, and whether or not the user may supply distributions for the variable, and marks variables selected for sensitivity analysis.

The *distributions* data set is where the user-supplied distributions are specified, except for those relating to specific scenarios. In general, the user may define multiple distributions for one SHEDS variable, along with conditions that indicate when each distribution applies. This allows for age, gender, or location-dependent distributions, among other options. However, to prevent the specification of incomplete or contradictory distributions, the interface only allows limited alteration; for full flexibility the user has to modify the Distributions data set directly, and must accept responsibility for ensuring proper definitions.

2.4 Chemical Usage and Concentrations

2.4.1 Contact Media

In SHEDS, humans (targets) are exposed to a chemical (agent) by coming into contact with various media. There are nine media considered in SHEDS, listed in Table 2-1. Two of these are air media and the chemical concentrations are reported in units of micrograms per cubic meter ($\mu g/m^3$, shown as ug/m3 in the code and printouts). The rest are surface media, with concentrations in units of micrograms per square centimeter (shown as ug/cm2). Version 4 allows the user to enter surface concentration distributions as either mass per unit area (ug/cm2) or mass per unit mass (ug/g) (except for pets, which only allows mass per unit area). The latter option is appropriate for measurements of soil or dust samples. SHEDS converts the (ug/g) values to (ug/cm2) using an effective dust or soil loading, randomly sampled once per person. Both forms (ug/cm2 and ug/g) can be entered, if desired, but note that the code will automatically sum them (after converting the latter units), so these should be used together only when they represent different chemical mass. Otherwise, double counting of the same chemical would occur.

Table 2-1. Contact media in SHEDS

Medium	Type	Description
Lawn	Surface	Grass and/or soil on lawns
Garden	Surface	Vegetable gardens
T_hard	Surface	Treated indoor hard (smooth) floors
T_soft	Surface	Treated indoor textured floors (e.g. carpets)
T_air	Air	Indoor air in treated areas
U_hard	Surface	Untreated indoor hard (smooth) floors
U_soft	Surface	Untreated indoor textured floors (e.g. carpets)
U_air	Air	Indoor air in untreated areas
Pet	Surface	Dog or cat fur

For surfaces, the concentrations represent only the portion of the chemical that is readily available for transfer; for example, the amount that could be picked up using a wipe sample or a PUF roller. This is less than the amount originally applied for various reasons; in particular, chemical penetrating deep into the material such as carpet is excluded. The surface loading will diminish over time, but in SHEDS it is not considered to be affected by human contact. In other words, it is assumed that repeated touches do not alter the surface loadings in the various media, beyond the typical loss rate. SHEDS calculates and reports the "chemfrac" variable, which represents the fraction of the available chemical that is transferred to the person, both on the diary event level and as daily totals (based on size of location such as house, garden, or lawn; the mass of chemical in the environment; and the mass transferred to the simulated person). This is included to ensure that conservation of mass is not being violated, per recommendation from the 2007 SAP. These fractions are usually a few parts to a few hundred parts per million. Any such losses are assumed to be included in the natural decline in surface concentrations over time (modeled using either the decay/dispersion method or the intervals method).

2.4.2 Application Scenarios

In SHEDS the chemical can enter the house through various chemical application scenarios, referred to as *scenarios*. These refer to combinations of chemical formulation, application method, and location. A SHEDS run may contain any number of scenarios. The default input files contain data for ten, but more can be added (or fewer used). In a cumulative or multi-chemical run, each scenario may contain one or more of the chemicals of interest. The names of the scenarios are arbitrary (up to 32 characters in length) and are descriptive only, and do not refer to (or limit) their properties in any way.

If the user enters concentration time series for the media directly into SHEDS (this is the *timeseries* option), then scenarios and their usage patterns are not relevant. The purpose of application scenarios is to aid in determining the chemical concentrations that are appropriate for each individual. If the Fixed Dates option is used, then the user must specify the scenario-date combinations on which usage is assumed to occur. Under Model-Determined Dates, each scenario has rules for determining its usage pattern. If desired, the user may specify co-occurrence rules which can alter the likelihood of scenario usage, based on when or whether another scenario was used by the same person.

The distinction between chemicals and scenarios should be noted. Even a single-chemical run may contain multiple scenarios (see Table 2-2). In a multi-chemical run, each scenario may contain one or more of the chemical agents. It is possible to have a single scenario that contains all the chemicals of interest.

Table 2-2. Application scenarios on the default SHEDS input files

Scenario	Description	Location	Requirement
Lawn_Granul	Lawn treatment (granular, push spreader)	Outdoor	Has_Lawn
Lawn_Liquid	Lawn treatment (sprayer or handwand)	Outdoor	Has_Lawn
Veg_Garden	Vegetable garden (dust, powder, or shaker)	Outdoor	Has_Garden
CC_Aerosol	Indoor crack & crevice treatment (aerosol)	Indoor	None
CC_Liquid	Indoor crack & crevice treatment (liquid)	Indoor	None
Ind_FIK	Indoor flying insect killer (aerosol)	Indoor	None
Ind_Fogger	Indoor fogger (broadcast)	Indoor	None
Pet_Spot	Pet treatment (spot-on)	Pet	Has_pet
Pet_Liquid	Pet treatment (liquid shampoo)	Pet	Has_pet

Scenarios are classified by their location of use: indoor, outdoor, or pet. Each person is assigned a re-entry time prohibition and a time-of-usage setting, for each of the three locations. The scenario location determines certain settings that apply; but in principle, SHEDS allows any scenario to increase the chemical concentrations in any (or all) media.

If the Fixed Dates option is used (see below), then all selected scenarios are used on the same dates by all persons in the model run. If the Model-Determined Dates option is used then the requirements must be met for a scenario to be possible. The user sets the probabilities p_lawn, p_garden, and p_pet for the population. The derived variables Has_lawn, Has_Garden, and Has_Pet will be each randomly set to either 1 (yes) or 0 (no), for each person.

2.4.3 Chemical usage patterns (applications)

Under the "Fixed Dates" option, the user specifies which dates and the hour of the day on which each scenario is applied. The dates are relative, with '1' being the first day of the simulation period. Multiple applications of the same scenario are allowed; these occur on different days but at the same hour. Multiple scenarios may be used on the same day, but each one may be used only once on a given day. For the Fixed Dates option, if a lawn application is selected, then automatically all simulated individuals are assumed to have a lawn. Similar logic applies to pets and to the vegetable garden, when applications of suitable type are specified.

Under the " Model-Determined Dates" option, SHEDS randomly generates chemical usage frequencies and dates according to user-specified rules. SHEDS does not contain any proprietary information, but users may supply any information as input data at their discretion. Each person may use one or more scenarios, or possibly none at all. The model implements the application scenarios only at the simulated individual's residence. The question of whether that person was the handler or applicator is determined separately. Since there is a (user-supplied) minimum age for being a handler, children typically would not be handlers, although they may be "users" (in the sense that they live at a place where the scenario was used). Like the Fixed Dates option, in Model-Determined Dates each scenario may be used on multiple dates, but not more than once on each day.

SHEDS does not request initial concentrations for the start of the simulation period to be input; nor does it assume that they are zero. Instead it generates them as a result of "prior usage," which refers to usage dates before the start of the simulation period. The concentrations resulting from such prior usage are modeled even though the human activities and the resulting exposure and dose are modeled only during the simulation period. In Fixed Dates, such "prior usage" is achieved simply by specifying usage dates that are less than 1 (which is always the first day of the simulation period). For example, usage on day (-7) would mean usage on the eighth day prior to the first simulation day. The chemical concentrations would therefore be reduced by a week of decay/dispersion by the time the simulation period started. If there was no prior usage, then the initial concentrations are set to zero by SHEDS.

Prior usage is determined slightly differently in the Model-Determined Dates option. First of all, the rules for determining usage apply at the annual level. For example, the frequency of use is specified as the number of days per year on which the given application scenario occurs (at the given house). This allows the user to alter the simulation period without having to alter all the input data referring to usage probability. Similarly, the monthly usage probabilities are expressed as fractions of all usage days occurring in a given month, for all 12 months, regardless of the simulation period. SHEDS then randomly generates usage dates for a full year (or multiple full years, if the simulation periods spans multiple years). Some of these usage days may occur before (or after) then simulation period. Any that occur after are ignored, as they cannot affect exposure during the simulation period. But those that occur before are modeled, to estimate any residual concentrations that may be present when the simulation period begins.

The user specifies the number of days prior to the beginning of the simulation period during which usage may be relevant. If the full year usage calendar does not contain enough such days, then the usage days from the end of the year are "wrapped" onto the prior year. For example, if the simulation period starts on January 15 and SHEDS is told to consider usage up to 30 days prior, then any usages during the last two weeks of December are copied onto the same days of the prior year.

2.4.3.1 Usage frequency

In the Model-Determined Dates option, several tests are needed to determine the usage dates. For each person, the code loops over all scenarios, considering the following questions: Does this scenario occur for this person? If so, then how often? On which specific dates does usage occur?

Some scenarios have a "requirement" or a prerequisite for use. For example, the two lawn scenarios require Has_lawn =1 (which represents “yes”), or else that person is automatically a non-user. This is determined on the OnePerson data set, based on the user input for p_Lawn. For example, if v1=0.75 on the record for p_Lawn on the Distributions data set, then each person has a 75% chance of being assigned as having a lawn (indicated by Has_Lawn=1 on the output file). Similar rules apply for the two pet application scenarios (which require Has_pet=1) and for the Vegetable Garden application (which requires Has_Garden=1).

Even when a person meets the specified requirement, a given scenario might or might not occur at the simulated person’s house. The SAS variable that controls this is called “p_User” and is scenario-specific. It represents the likelihood of usage, given that the appropriate requirement has been met. Thus, in the above example, if p_User = 0.50 for granular lawn treatment, this would mean that about 37.5% of the individuals in the model run would have granular lawn applications (75% have lawns, and half of those treat them in this manner) at least once per year. However, for shorter simulations many of these people would not use the granular lawn scenario during the simulation period. If the modeler wished to confine the model run to users, but still wanted the usage dates to vary, then Model-Determined Dates with p_Lawn=1 and p_User=1 should be selected.

For the scenarios that occur, the number of usage dates (number of applications) is determined next. As mentioned earlier, this number is on an annual basis. If the simulation period covers multiple years, then the number of usages is scaled up. A separate draw of usage frequency is made for each scenario.

2.4.3.2 Usage dates and co-occurrence

Once the usage frequency is known, the specific application dates must be determined. The starting basis for this comes from the ‘Month’ and ‘DayOfWeek’ probability vectors. Every possible day on the extended application calendar is assigned a probability given by the product of the Month probability and DayOfWeek probability corresponding to that date. For example, Monday January 1 is assigned the product of the Monday probability and the January probability. When all possible days are assigned in this manner, the adjustments for co-occurrence with other scenarios are made, based on the values of the influence factors and influence widths, discussed below, Figure 2-2 illustrates the selection of dates for the first application scenario. Figure 2-3 displays how the model chooses dates for a second application scenario.

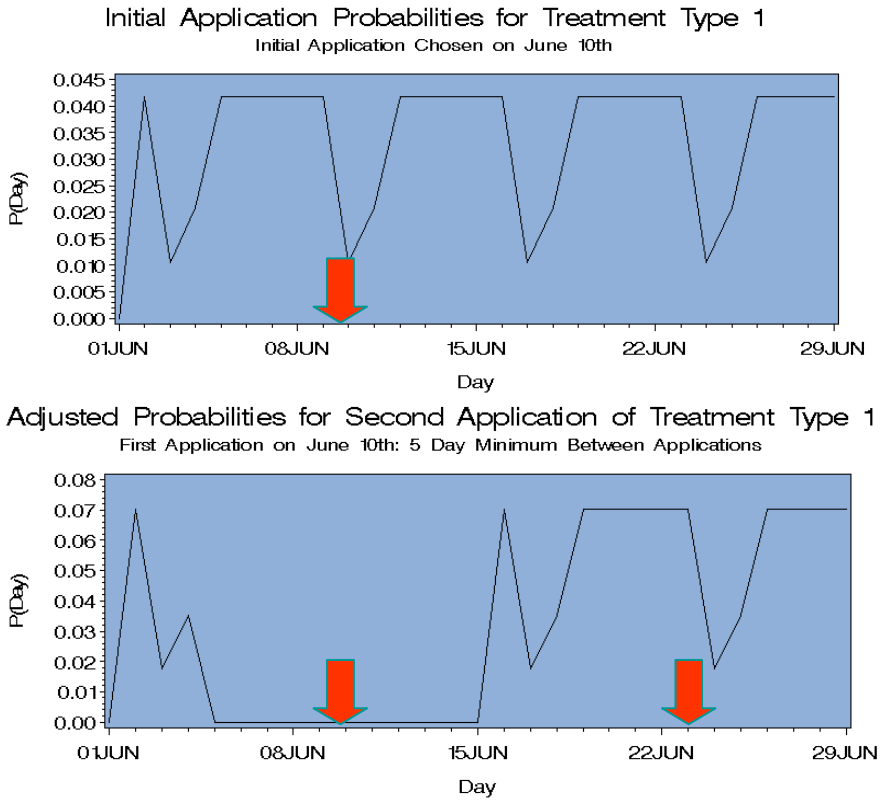
The adjustments are calculated on a single pass over the scenarios rather than iteratively, so the adjustments take into consideration only the dates that have already been assigned to this person. The scenarios are each assigned a priority by the user. The first scenario on the list never has adjustments for others, although if it is used multiple times then adjustments for blackout periods around already-selected dates may apply. The second scenario may have adjustments that depend on the dates when the first was used. The third can be adjusted for either of the first two, and so on. The probability adjustment takes the form of a multiplier (influence factor) that is applied to all days within the specified number of days (influence width) of an application of the indicated scenario. For example, suppose scenario #1 is used on June 10, scenario #2 is used on June 20, scenario #3 is used on June 30; and that the user requests that scenario #4 is only half (i. e., influence factor=0.5) as likely to be used within 7 days of either #1 or #2

(or both). Specifically, all dates from June 4 through June 26 inclusive have their probabilities halved when choosing usage dates for scenario #4. The dates around June 30 are unaffected because scenario #3 was not included in the list that affects #4.

After these probability adjustments are made, the total probability over all days is calculated, and each individual daily probability is divided by this total. This ensures that the sum over all days equals one. One date is selected at random from this probability vector covering the full year, which becomes a usage date. If this scenario has multiple usages, the remaining dates would still have to be determined. Suppose the first date for #4 was found to be September 15 and the blackout period is 14 days. This means that no two usages may occur less than 14 days apart. Therefore, the probabilities on all dates from September 2 through September 28, inclusive, must be set to zero. September 1 and September 29 are not affected, as they are each 14 days away from the application date in question. The minimum blackout period is 1, which means that two applications of the same type cannot occur on the same day. Once the blackout period is set, the probability vector is again normalized by summing the probabilities over all days and dividing each one by the total. Then another date is selected. This continues until the correct number of usages have been assigned, or all possible dates have been blacked out. In the latter case, that scenario does not occur again, so the number realized is less than the number specified by the NumApps value. Note that the usage dates are not drawn in chronological order; the first one assigned is not necessarily the first on the calendar.

For each application scenario the user specifies:

- Evaluation order
- Usage probabilities by month and day of week
- Probabilities based on number of applications
- Minimum number of days between applications

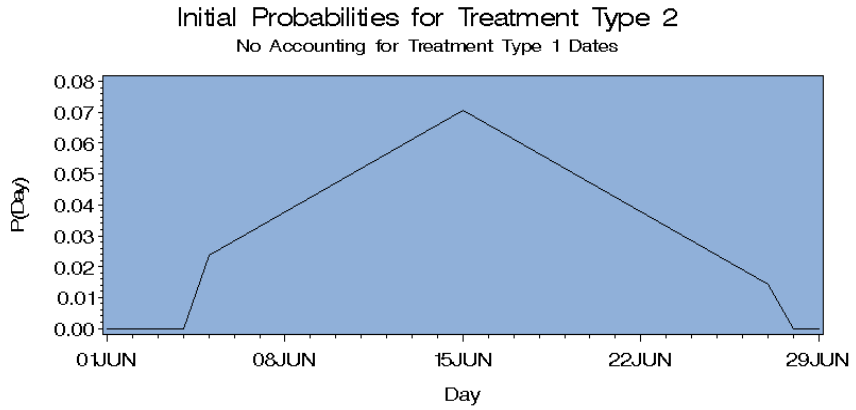


Initial probabilities for scenario 1. First usage date randomly chosen is June 10th.

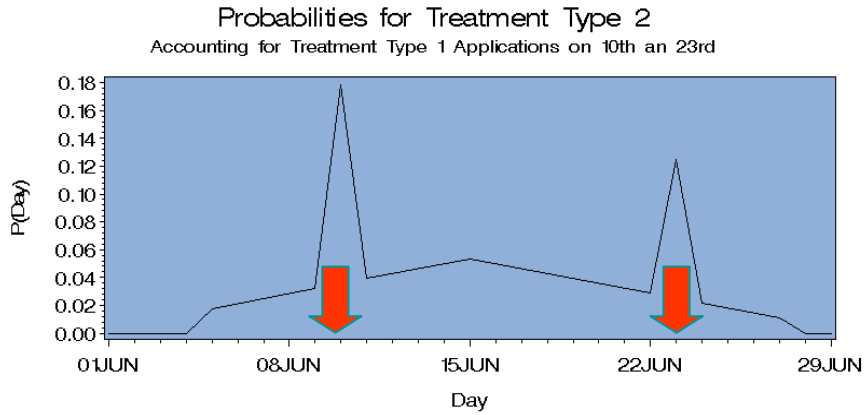
Probabilities adjusted for June 10th application date assuming a 5 day minimum time between applications. Second usage date chosen is June 23rd.

Figure 2-2. Choosing usage dates for a single application scenario.

- Co-occurrence relationships among application scenarios
- User selects which other scenarios affect usage of current scenario
- Specify influence factor and influence width for each pair of related scenarios



Initial probabilities for scenario 2 not accounting for usage of scenario 1.



Probabilities adjusted for usage of scenario 1 on June 10th and 23rd. In this case the adjustments are upward, increasing the likelihood that scenario 2 will be used on either June 10th or June 23rd.

Figure 2-3. Co-occurrence between scenarios: choosing usage dates for scenario 2.

Each usage is assigned to a specific hour of the day, based on sampling the v_TimeOfUse probability vectors. There are three such vectors, one for indoor scenarios, one for outdoor, and one for pets. The v_TimeOfUse vectors are offset by seven hours, so the first element refers to the probability of usage at 7 a.m., with each subsequent element corresponding to another hour delay.

2.4.3.3 Re-entry prohibition times

Often, people may choose to avoid (or are told to avoid according to pesticide product labels) entering a treated area shortly after a chemical application is made. In SHEDS, this is called the re-entry time prohibition – also referred to in SHEDS as the "ban" period. SHEDS does not actually alter the locations on the activity diary, it merely sets the contact time with treated areas to zero. Thus, any in-home diary events during the ban period are spent in the untreated part of the house, and any out-home events are spent away from the lawn and garden.

Three variables are used: one each for indoor, outdoor, and pet scenarios. Unlike version 3, any type of allowed SHEDS distribution may be used for the re-entry time, and the result is rounded down to whole hours. Furthermore, contact with chemical does not necessarily resume immediately when the ban period ends. The activity diary must also place the person in the affected location. For example, if the home is treated at 9 a.m. and a 2 hour ban is in effect, a child might not contact the treated area until they come home from school several hours later. The ban is only a lower bound for the time until contact occurs.

2.4.4 Methods for determining media concentrations

SHEDS version 4 has three general options for generating exposure concentration time series for the various residential contact media: decay/dispersion, intervals, and user-supplied time series. These options are discussed below. A given SHEDS model run may use only one of the three methods, so the same method must be used for all contact media.

However the residential media concentrations are determined, SHEDS produces an hourly concentration time series for each medium, covering the full simulation period. For example, a 365-day simulation has 8,760 hourly concentration values. A new feature in version 4 is that the simulated person is now allowed to move within a single location during a diary event, so the concentration encountered on that event is a duration-weighted average of the media in that location. Each of the three methods of producing the hourly concentration time series is discussed in the following sections.

2.4.4.1 Decay/Dispersion Method

The decay/dispersion method uses first-order decay logic for concentrations in treated areas, and uses a dispersion mechanism to move the chemical from treated to untreated areas indoors. The treated media have sudden jumps in concentration at the application dates and times, and chemical levels gradually decline thereafter. The decline is modeled as a piecewise linear approximation to an exponential decay process. This method requires distributions for the initial concentrations in each treated medium, and the decay rate. If there are pre-existing concentrations in these media, they are added to the initial concentrations from the new usage event.

SHEDS version 4 uses a single decay rate for each scenario, apart from the air concentration during the first 24 hours after an indoor application. This is supported by runs of a fugacity model being developed for SHEDS (Glen et al., 2007), which show very similar decay rates in all media. The reason

is that the chemical will redistribute itself to reduce large fugacity differences, which means that the concentration ratios in the various media tend to stabilize. Basically, all media decay at the rate of the one that is naturally slowest. For example, indoor air is exchanged with the outdoors several times per day, which should imply a very high decay rate (more than 99% per day). But the air quickly reaches a lower fugacity than the carpet and hard floor compartments, so chemical flows from the floor into the air until balance is restored. Thus, the typical drop in air concentration from one day to the next is nearly the same as the drop in the floor concentration, meaning that the decay rates are nearly equal.

The "decay rate" in SHEDS is input as a fraction lost per day. A decay rate of 0.15 implies that the concentration drops by 15% of its current level each day. Since the current level declines over time (in the absence of new applications), the concentration time series flattens out as it approaches zero, approximating an exponential decay. The model converts the daily decay rate D_d to an hourly decay rate D_h, based on the relationship $(1-D_d) = (1-D_h)^{24}$. That is, SHEDS calculates the hourly decay rate needed to result in the daily loss fraction specified by the user. If the daily fraction lost is much less than one, then the hourly fraction lost is close to 1/24 of the daily amount, but when losses are large this changes. Each hour, the concentration in a treated area is given by the value for the prior hour times $(1-D_h)$, unless a new usage occurs. Equation 2-1 expresses this relationship mathematically:

Equation 2-1

$$\mathbf{C_t = C_{t-1} (1 - D_h) + A_t}$$

where C_t = concentration at hour 't'

C_{t-1} = concentration at hour 't-1'

D_h = fractional decay rate (constant over time series), $0<D_h<1$

A_t = additional concentration from new application at time 't' (if any).

Indoors, the untreated media slowly accumulate chemical from the treated area. The modeling of this process is slightly different in version 4 than in version 3, although the results are similar. Draft fugacity model runs have shown that the untreated-to-treated concentration ratio behaves differently in air and in surface media. In air, the ratio quickly (in a few hours) approaches a constant value and thereafter does not change, assuming that the building ventilation properties remain fixed. The ratio depends largely on the rate of air exchange between the treated and untreated areas and on the indoor-outdoor air exchange rate; these are competing processes, with one raising the concentration in the untreated area and the other lowering it. In SHEDS, this ratio is taken to be constant at all times, although it may vary from person to person (that is, from one house to another).

For indoor surface media, in the absence of new usage events, the untreated-to-treated concentration ratio was found in the draft fugacity model to increase almost linearly with time until it approaches one. This rule is made exact in SHEDS-Residential version 4.0. As mathematical functions of time, the treated room concentration T(t) and the untreated room concentration U(t) are taken to be:

$$T(t) = T_0 \exp(-k\, t)$$

$$U(t) = f\, t\, T_0 \exp(-k\, t)$$

This function U(t) starts at zero, rises to a maximum at a time t=1/k, and thereafter declines gradually to zero. The user supplies the untreated-to-treated concentration ratio at the time when the untreated area reaches maximum concentration. From the above equations, this ratio must be (f/k). If time t is expressed in hours, then k is the hourly decay rate. Hence the constant 'f' is the product of the user-supplied ratio at untreated maximum with the hourly decay rate. The other constant T_0 is simply the initial concentration after usage.

The above mathematical functions are not correct at very large times, because eventually U(t) will exceed T(t) and the difference will continue to increase. In practice, if that ever occurred then the roles of treated and untreated areas would reverse and the net flow would be in the other direction. So the rule is imposed that U(t) cannot exceed T(t). The modified equation for U(t) is:

$$U(t) = T(t) * \min (f\ t, 1)$$

The min function returns the lesser of its arguments, so U(t) equals the form given earlier until time t = 1/f, and thereafter U(t)=T(t). This equation is applied separately to the carpet and the hard floor, although only T_0 differs between the two. Typical values for the inputs would be around 0.1 for the untreated-to-treated concentration ratio at the untreated maximum, and k = 0.005 hr^{-1}. This gives f=2000 hr or about 83 days. By this time, both the treated and untreated concentrations have dropped to negligible levels. Since the concentration time series are hourly, t increases by one on each step, and it follows that

$$U(t) / T(t) = U(t-1) / T(t-1) + f\ ,$$

until the ratio equals one. This last form is the version that appears in the SHEDS code.

When a new usage event occurs, the treated concentration T immediately increases by the media-dependent amount A_t as shown in Equation 2-1. The untreated concentrations U are not affected initially. Indoors, three media are affected (air, hard floor, and carpet), and the same percentile is drawn from the three distributions to determine the increases in concentration. If handler exposures are modeled, then this same percentile is also drawn from the application rate distribution. SHEDS adds the new initial amount T_0 to the current treated concentration, and recomputes the U(t)/T(t) ratio. This new ratio is incremented by an amount 'f' each hour thereafter.

Based on the fugacity model being developed for SHEDS and available measurements data, during the first 24 hours after an application, the air concentration decays very rapidly and at a non-constant decay rate. Initially, the decay rate is driven by the deposition of large particles onto the floor. As these are lost, the average particle size in the air decreases, and smaller particles deposit more slowly. Later on, the main loss is air exchange with the outdoors. However, if the air is still at a higher fugacity than the floor, resuspension of particles on the floor will not be able to replace these losses. Eventually, the air concentration will drop to a point where chemical flows from the other compartments will largely balance the losses due to air exchange. After this, the air will decay at virtually the same rate as the other compartments.

Typical air exchange rates for houses are around one exchange per hour. At that rate, without any replenishment, the air concentration would be reduced by a factor of Exp(-24) in one day. In practice, air concentrations typically drop by a factor of between one hundred and one thousand on the first day, which is a far smaller reduction than air exchange alone would produce, indicating that significant replenishment occurs before the first day has passed. Thus, the transition to a common decay rate with the other media occurs within the first day.

Data from Keenan (2007) show that the relationship between the logarithm of air concentration and the logarithm of time is very close to linear, with the R^2 for a linear regression of the pyrethroid air concentration at 1 foot elevation being 0.988 for Log(c) versus Log(t). The user supplies two air concentrations as inputs for indoor scenarios: one for the first hour after application and the other after 24 hours. SHEDS interpolates between these two concentrations at intermediate times, assuming a linear relationship between Log(c) and Log(t). The user enters the overall decay rate for all compartments, but does not enter any other decay rate especially for air.

Outdoor and pet scenarios are simpler than indoor because there is a single affected medium and no dispersion calculations are needed. The concentration follows the same decay logic as is used for indoor treated media, although the decay rates may be different. The outdoor and pet scenarios are assumed to result in negligible air concentration (except during application).

Both indoors and outdoors, it is likely that multiple scenarios will affect the same media. In such cases, separate concentration time series are calculated for each. A new feature in version 4 is that the fraction of the house that is treated is scenario-specific. It is possible to be in a part of the house that is treated under some scenarios but untreated under others. This point is addressed in the section below on duration-weighted concentration averaging.

2.4.4.2 Intervals Method

An alternative to the decay/dispersion approach is the intervals method, which requires that four concentration distributions be entered for each contact medium, specifically for the time periods <1 day, 1-7 days, 8-30 days, and 31 or more days after the most recent usage. In some practical measurement situations, this sort of input data might be more readily available than the decay/dispersion inputs.

As with the decay/dispersion method, each scenario is analyzed separately in the intervals method. Hourly time series for the entire simulation period are constructed for each scenario. Hence, the time intervals (such as 12 days since the last fogger usage, for example) are scenario-specific and are unaffected by the use of other application scenarios.

When usage occurs, a percentile is chosen at random. Small percentiles correspond to light usages of the chemical, while larger percentiles represent heavier usages. On the usage day, from the hour of application until midnight, the corresponding percentile of the concentration distribution for <1 day is used as the concentration. For the next seven days, the same percentile of the 1-7 day distribution is used. Later, the same percentiles of the 8-30 day and 31+ day distributions are used. When another application occurs, the process starts over. The only exceptions are that the concentration cannot rise when going from one time period to the next without a new usage event, and the concentrations cannot fall as a result of usage. If either exception occurs, then the prior concentration continues to be used.

With multiple occurrences of the same application scenario, the intervals method only considers the most recent one, unlike the decay/dispersion method. This was deemed appropriate since measured concentrations (used as a source for input distributions) would presumably already include any effects of multiple applications.

The intervals method requires the user to supply a concentration distribution for each of the 4 time intervals, for each treated medium. It also asks for untreated/treated concentration ratios for each of the 3 indoor media types (air, hard floor, carpet), for each interval. If the user does not wish to include dispersion to untreated areas, then these ratios should be set to point values of zero.

2.4.4.3 User-Specified Concentration Time Series

This is the simplest option for determining the concentration time series. The user supplies a file (a SAS data set) containing daily time series for each contact medium. This data set has one record per day. The list of variables must include one called “date” which indicates the date for each record, using SAS date format. The other required variable is "chemical", which contains the chemical name. If the model run contains multiple chemicals, then there is one record for each chemical-date combination. It does not matter whether the sort order is chemical-date or date-chemical. The other variables on each record have names in the form “medium_house#” and contain chemical concentrations. For example, if there were data from three houses, then the treated air concentrations have variable names “Tair_1”, “Tair_2”, and “Tair_3”. The corresponding carpet concentrations would be “Tsoft_1”, “Tsoft_2”, and “Tsoft_3”. The units do not appear on this file, and the user must ensure that they match the SHEDS requirements: ($\mu g/m^3$) for air, and ($\mu g/cm^2$) for surfaces.

The numerical suffix is called the "house number." SHEDS selects one house at random for each person, and uses all the concentrations from that house together. This allows for the use of sets of measurements that belong together. It is reasonable that high carpet concentrations coexist with high hard floor concentrations in the same house. Without house numbers on the input file, SHEDS might randomly draw very high carpet and very low hard floor concentrations for the same person, for example. The data on the input file do not have to come from actual house measurements; they could be modeled or constructed as hypothetical examples. But even so, the user can control which examples of each medium are to be used together.

In the time series option, the entire house is assumed to be treated, as there is no usage information to determine otherwise. The untreated media are not used. Other variables may appear on the input data set apart from the ones listed above; they are not read by SHEDS. It might be useful to have a variable for comments or notes to help document the data set.

The houses do not need to use the same start and stop dates, but each house must have no gaps (days with missing data) between the start and stop dates. All of the media for any given house must have the same start and stop dates. As an example, suppose there are just three houses and two media with non-zero concentrations. The following input file shown in Table 2-3 would be valid:

Table 2-3. Example input file for concentration time series.

Chemical	Date	Thard_1	Tsoft_1	Thard_2	Tsoft_2	Thard_3	Tsoft_3
X	19JUL2010	.	.	1.14	0.83	.	.
X	20JUL2010	.	.	1.13	0.81	.	.
X	21JUL2010	0.75	0.56	1.11	0.78	.	.
X	22JUL2010	0.71	0.53	1.10	0.77	.	.
X	23JUL2010	0.69	0.51	1.08	0.75	0.99	0.62
X	24JUL2010	0.67	0.49	.	.	0.96	0.60
X	25JUL2010	0.64	0.46	.	.	0.93	0.58
X	26JUL2010	0.62	0.44	.	.	0.91	0.57
X	27JUL2010	.	.	.	.	0.88	0.55
X	28JUL2010	.	.	.	.	0.86	0.53

In this example, Tair and Tdust are missing from the file, so they will be assumed to be zero at all times. Missing data in SAS are indicated by periods; none of these houses have data covering all 10 days. The dates are SAS dates; these appear as shown when viewed or printed, but are actually stored as integers. If these are to be created or modified, do not replace them with explicit character strings. SAS

will not equate the 9-character string '19JUL2010' with the date 19JUL2010; instead the integer 18462 must be entered for this date. There are built-in SAS date functions that perform such conversions.

Each person simulated will be assigned to one of the three houses at random. The simulation period for that individual is then set according to the start and stop dates for that house, provided they fall within the start and stop dates of the simulation period as a whole.

The user does not need to supply data for more than one house. If the input file contains data for N houses, then roughly (1/N) of the people will be randomly assigned to each house. For each simulated person, all the concentrations are taken from the same house. After the model run is complete, the user can examine the exposure distribution within houses, between houses, or overall. This may provide information on the contribution of variance in concentration to the variability seen in exposure or dose.

When defining a SHEDS-Residential job, the user specifies a start date and a number of days, defining the simulation period. For the FixedDates and VariableDates options, all simulated persons share the same simulation period. For the TimeSeries option, these dates are used to subset the input file. The dates on the timeseries input file may begin before and/or extend after the dates for the simulation period, but only the dates within the simulation period are utilized.

For example, suppose the start date is set to July 25, 2010 and the number of days to 7. Then only the last four records in Table 2-3 would qualify as belonging to the simulation period. Since there is no valid data remaining from house #2, it is removed as a possible candidate, so half of the simulated persons would be assigned to house #1, and the other half to house #3. In this case the persons assigned to house #1 would have 2-day exposure simulation periods, while those in house #3 would have 4-day simulation periods.

2.5 Generating Longitudinal Activity Patterns for a Simulated Individual

2.5.1 Selecting Activity Diaries

The residential activities for each SHEDS individual are determined by selecting a set of one-day sequential time-location-activity diaries from surveys contained in EPA's CHAD (Consolidated Human Activity Database; http://www.epa.gov/chadnet1; McCurdy et al., 2000). These diaries are matched to the SHEDS individual by age-gender cohort, and by season and day-type (i.e., weekend or weekday). The CHAD database has been augmented since the one used in SHEDS version 3, and now contains around 34,000 diary days.

The following studies in CHAD provide diaries for young children: two waves of the Panel Study on Income Dynamics conducted by the University of Michigan; two versions (air and water) of the National Human Activity Pattern Survey; the California children's study; the Cincinnati study, and the Ozone Averting Behavior study. Three studies focus only on older school-age children (the Los Angeles elementary and high-school studies, and the California youth study). One study focuses on the elderly (the Baltimore study), while the others include the general adult population. These include the Denver, Washington, and Valdez studies, the California adults' study, most of the diaries in the Cincinnati study, the "EPA" diaries, and most of the NHAPS study. Appendix D contains summary information and diary counts for the CHAD diaries used in SHEDS. It also contains an excerpt of a CHAD diary and listings of all CHAD activities and locations.

Age and gender variables are used to subset the activity diaries because they are important predictors for time spent outdoors (Graham and McCurdy, 2003). Within the age-gender cohorts, the

diaries are divided by SHEDS into eight subgroups (weekday and weekend within each of the four seasons). Day of week (i.e., weekend or weekday) and season of year are the most important variables for compiling a longitudinal diary. Other variables were not used because only a few are common across all CHAD surveys and because finer divisions would reduce the number of diaries available for random sampling. There are 10 age groupings as described above (not counting those less than 1 year, which are not currently modeled in SHEDS), 2 genders, and 8 subgroups in each, creating 160 diary pools.

A longitudinal activity diary in SHEDS is one that covers the entire simulation period (typically one year or less). It should be unique to each simulated individual. Since most people in the CHAD database supplied just one day of diary data, a method is needed for combining many CHAD diaries to cover a longer period. SHEDS version 4 has two such methods; the user decides which method to use (the same method is used for all persons in a given model run).

One option is the eight-diary method based on Xue et al., 2004, which was also used in SHEDS-Residential version 3. (See Figure 2-4.) For each person, 8 CHAD diaries are selected, one weekend and one weekday diary from each of 4 seasons. These are then assembled into a single long diary by repeating the appropriate one, according to the calendar. Thus, if the year starts on a Sunday, one copy of the winter weekend diary is used for the first day, followed by five copies of the winter weekday diary, then two more copies of the winter weekend diary, and so on until the next season starts.

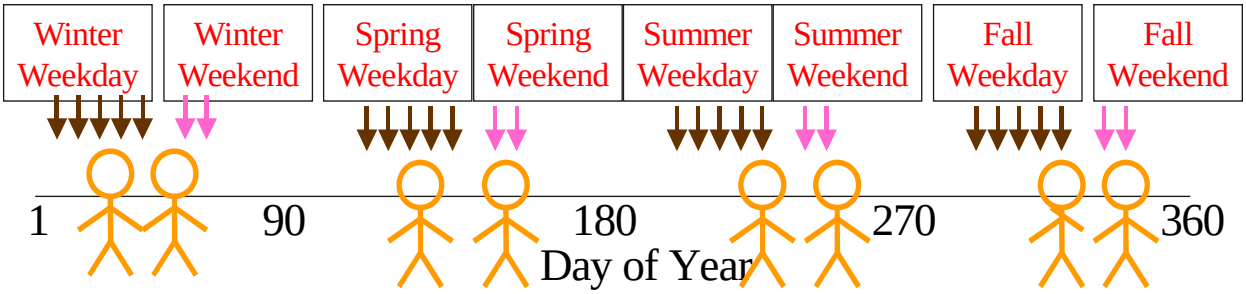


Figure 2-4. SHEDS eight-diary approach for simulating longitudinal activity patterns.

For assembling longitudinal diaries, Xue et al. (2004) suggest that eight days per year spread evenly across seasons is a reasonable number to use in that it captures most of the relationship between intra- and inter-personal variability with respect to daily time spent outdoors, based on a Southern California study of 160 children. It is important to note that while the same CHAD diary is used repeatedly within a year, this does not mean that the times spent in contact with treated media are the same from day to day. As explained below (see “Determination of Contact Time” section), the potential contact time (the time spent in locations with chemical present) is randomly divided into time in treated or untreated areas, on the various surfaces (carpet, hard floor, lawn, garden, or other). Thus, the amount of contact with each medium will vary from day to day, even if the same CHAD diary is used.

The other diary assembly option is discussed in detail in Glen et al. (2008). It is based on making a new diary selection each day, but altering the selection probabilities to induce set targets for the within-person variance, between-person variance, and day-to-day autocorrelation in a selected key variable from the diaries. The user selects the key variable from a pre-defined list, and specifies the D (diversity) and A (autocorrelation) statistics. "D" is the ratio of the between-person variance to the total variance in the across-person rankings of the key variable, and is defined only on the population as a whole. "A" is the population mean of the autocorrelation in the within-person rankings of the key variable.

The new method potentially selects a different diary every day (and would if the diary database were sufficiently large). However, often there are not enough distinct diaries, so the same ones are

usually selected multiple times. The D statistic controls how similar all the diaries are that are assigned to a given person. As D approaches zero, all diaries in the appropriate pool are selected at random with equal probability. As D approaches one, just one diary in each pool would have a non-zero selection probability, although for different persons this diary would likely be different. Neither extreme is expected to be realistic in general. For example, some D values based on outdoor time as the key variable have ranged from 0.1 to 0.5 (Glen et al. (2008), Isaacs et al. (2007), Isaacs et al. (2009)). More longitudinal data are needed to refine the D and A statistics, especially for the dietary algorithms.

The D statistic determines the selection probabilities for each diary and therefore affects the similarity of the chosen diaries assigned to each person. However, it does not affect the order in which these diaries are "strung together." The A statistic controls the similarity of one day's diary to the next subject to the constraint already imposed by D. High A values (near 1) imply a strong positive relationship, meaning that (for example) the day with the highest key variable score (such as the most outdoor time) will tend to be surrounded by other days with high scores. An A value of zero indicates that the days are independent. Negative A values imply a tendency toward opposites, so a high day is likely followed by a low day, and vice versa.

The new D&A method may use any day-type pooling, and has been set up to use the same season and weekday-weekend pools as are used by the eight-diary method. Therefore, the processing and pooling of the CHAD database applies to both diary assembly methods and does not have to be altered when the assembly method is changed.

2.5.2 Determination of microenvironment and diary details

The longitudinal diary assembled from the one-day CHAD diaries gives the general activity, location, and duration information that characterize the time-activity behavior of the simulated individual. The CHAD diary events range in duration from 1 minute to 1 hour, and never cross an hour boundary. Hence, each clock hour contains one or more complete CHAD diary events. Many random variables are resampled hourly in SHEDS, and it is convenient that the properties determined by these random variables remain fixed over the duration of a diary event.

Every diary event in CHAD has a specific start time, duration, and stop time, all of which are in whole minutes. There are no time gaps and no time overlaps (with two or more simultaneous events for the same person). Occasionally, the original study had a time gap or a partial description of an event. In such cases, CHAD creates an event to fill in the gap, but the activity and/or location codes are set to missing values (either "U" or "X" in CHAD). By default, SHEDS rejects any CHAD diaries that have one hour or more of missing location, or 2 hours or more of missing activity codes. These thresholds can be altered on the DiaryQArules input file, but if this is done, then the DiaryPreProcess module needs to be run again. Normally, a SHEDS run skips this step since it has already been done for the default input files, and only needs to be repeated if the database changes or the diary rejection rules change.

When a missing activity or location is encountered in SHEDS, the simple rule is used of assuming the prior activity or location still applies. This is invoked after the longitudinal diary has been assembled, so occasionally the prior activity or location may come from the prior day, possibly from a different CHAD diary. SHEDS groups the CHAD events into the following six microenvironments ("micros"): in_home, out_home, in_other, out_other, in_vehicle, and sleep. The "sleep" micro is based on activity, while the others are determined by the location of the diary event. The mapping from CHAD location codes to SHEDS micros is in the "diarylocs" input file. This file can be edited in SAS to change the mapping, but not directly via the SHEDS interface.

In addition to determining the microenvironment, the diary detailing module in SHEDS performs the following tasks: (1) determining the extra bathing (since CHAD underreports bathing) events; (2)

determining hand washing events; and (3) marking sleeping events. All three are marked using the "special" variable: "Z" indicates sleeping, "B" a bath or shower, and "W" marks hand washing. Sleeping not only excludes the possibility of bathing or hand washing, but also prevents all new exposure except for inhalation. Bathing and hand washing may co-exist with other activities.

"Bath events" in SHEDS include showers. Bathing is underreported in CHAD because the database is a compilation of many activity studies, some of which did not report bathing as a separate activity. In some cases it is subsumed under a broader CHAD category such as "personal care," but one cannot infer that all personal care includes a bath, as this category is often used multiple times per day. However, other studies in CHAD have explicit bath/shower events. The rule in SHEDS is that there is a maximum allowed interval between baths, which is one to seven days. The default data show that about 75% of persons take a daily bath/shower. Each person in SHEDS is randomly assigned a maximum interval based on this distribution. Then if the assembled longitudinal diary has periods longer than this maximum, extra bathing events are forced (that is, the setting special="B" is added to the diary event) when the maximum interval is reached. This has implications in the exposure module.

Hand washing is never recorded in CHAD, because the duration is too short. Some of the studies in CHAD imposed a minimum 15 minute duration on diary events; but even those that did not tended to underreport events of less than 15 minutes duration. Since hand washing affects the dermal loading in SHEDS, this information is stochastically generated and added to the diary. Each person is assigned a value for HandWash_mean, which is the long-term average number of hand-washings per day. This is translated by the code into an hourly probability. A random check is made once per hour against this probability. Hence, it would be more exact to say that HandWash_mean is the average number of hours each day during which hand washing occurs. Two or more actual hand washings during the same clock hour would only count as one in SHEDS.

2.6 Contact Duration and Concentration averaging

While the activity diaries provide the sequence of micros visited by each simulated person, the diaries do not contain enough detail to determine the amount of contact with various media. SHEDS therefore stochastically generates contact durations for the various media for the diary events in potential exposure locations. The relevant contact media in each micro are listed in

Table 2-4:

Table 2-4. Micros and associated contact media

Micro	Description	Media that affect exposure
in_home	inside residence (and awake)	carpet, hard floor, indoor air, pet
sleep	sleeping at home	untreated air
out_home	outside around home	lawn, garden, pet
in_veh	in a vehicle	none
in_oth	inside another building	none
out_oth	outside away from home	none

In SHEDS the carpet, hard floor, and indoor air microenvironments each consist of two time series for each scenario, since the house is divided into a "treated area" and an "untreated area." When a chemical is used in the home, the untreated area is initially unaffected, but gradually chemical builds up there by dispersion from nearby treated areas. Eventually, the treated area concentration drops to the point where the supply of chemical to the untreated area is not sufficient to maintain the concentration, and the untreated area concentration thereafter declines as well.

The SHEDS code is modularized, and two of the most prominent pieces are the Concentration Module and the Exposure Module. In the former, hourly time series of chemical concentrations are prepared for each contact medium, for each scenario (when application scenarios are relevant). With 10 scenarios and 9 contact media, there would be 90 such time series. These are recalculated for each simulated person since their chemical usage patterns and house characteristics may differ. But as seen in Table 2-4, the Exposure module only requires 7 different media-specific chemical concentration time series. It does not matter to the exposure calculation which scenarios provide the concentration; only the total is relevant. Another complication in SHEDS is that the treated areas may be scenario-specific. For example, if a fogger has a larger treatment area than a crack-crevice application, then it is possible that the simulated person may contact a surface that is considered to be treated by the fogger, but untreated by the crack-crevice application. Therefore, simply summing all the treated media concentrations and summing all the untreated media concentrations would not be sufficient to determine all the total concentrations that could be encountered.

SHEDS solves the above difficulties using an approach called "duration weighted average concentrations." This involves determining the time fractions of each diary event that are spent with each contact medium, and then taking appropriate time-weighted averages. The basic principle is that the products of (concentration x time) are held constant. For example, suppose a one-hour event consisted of spending 40 minutes in a treated area and 20 minutes in an untreated area. From the standpoint of the exposure equations, that is equivalent to spending one hour in contact with an average concentration consisting of 2/3 of the treated value plus 1/3 of the untreated value. Using such averaged concentrations, the results could be summed across scenarios. Any differences in the treated areas have already been accounted for in the form of scenario-specific weights used to obtain the averaged concentrations. These computations are performed in the Concentration Module. Seven hourly concentration time series are produced, one each for carpet, hard floor, indoor air, sleep, lawn, garden, and pet.

2.6.1 The relationship between scenarios, locations, contact media, and micros

In SHEDS, chemical usage is tied to scenarios. There are 9 default scenarios listed in Table 2-2, but both the number and the names of the scenarios are data-driven and can be changed by editing the input files. There are no references in the SHEDS code to any specific scenarios; the locations, contact media, and micros are fixed lists and are referenced directly in the code. Each scenario is assigned a particular location, which determines which contact media may be affected and when these media may be contacted. The three locations are Indoor, Outdoor, and Pet. The relationships are shown in Table 2-5:

Table 2-5. Associated locations, micros, and contact media

Location	Micro	Contact Medium
Indoor	In-home	Treated Hard Floors (T_hard)
Indoor	In-home	Treated Soft Floors (T_soft)
Indoor	In-home	Treated Air (T_air)
Indoor	In-home	Untreated Hard Floors (U_hard)
Indoor	In-home	Untreated Soft Floors (U_soft)
Indoor	In-home or sleep	Untreated Air (U_air)
Outdoor	Out-home	Lawn
Outdoor	Out-home	Garden
Pet	In-home or out-home	Pet Fur

The activity diaries discussed in section 2.5 provide the sequence of micros, as each CHAD event is assigned to a specific micro and has a definite start time, duration, and stop time. However, the times spent on the various contact media are not uniquely determined by the diaries. For example, any in-home diary event may be partitioned into four durations that sum to the total CHAD event duration: on treated carpet, treated hard floors, untreated carpet, and untreated hard floors. Similarly, out-home events are partitioned into three partial event durations: on lawns, in gardens, and neither. In version 4, time spent in contact with pets is considered to be in addition to contact with other surfaces. For example, a one-hour indoor event might entail 40 minutes on treated carpet and 20 minutes on treated hard floors (and therefore no time in untreated areas), but may also include anywhere from zero to sixty minutes of pet contact. For out-home events, it is assumed that the entire lawn is treated (if it is treated at all) and similarly for the garden, but it is possible to spend out-home time on neither (for example, playing basketball on the driveway). Note that partial lawn treatments can be represented in SHEDS by adjusting the user-specified input for lawn size, to reflect only the treated portion. This "neither" category is similar to the indoor untreated surfaces, except that SHEDS assumes that no chemical migrates to those areas and hence no new exposure can occur there.

2.6.2 Input variables controlling contact duration

There are five inputs to SHEDS that affect the contact duration, as listed in Table 2-6:

Table 2-6: SHEDS input variables that affect contact duration

Input variable	Description
f_carpet	Average fraction of in-home time spent on carpet
f_lawn	Average fraction of out-home time spent on lawn
f_garden	Average fraction of out-home time spent in garden
f_pet	Average fraction of at-home time (in-home or out-home) spent around pet
f_AreaTreated	Average fraction of in-home time spent in treated area (for each indoor scenario)

The last of these is scenario-specific, so with five default indoor scenarios there are five inputs for f_AreaTreated. The others apply to all scenarios affecting those media. Distributions may be entered for all these variables. SHEDS samples each of these variables once per person (which means that

f_AreaTreated is sampled once per scenario, per person). For example, suppose that the average home is 60% carpeted. Specific homes may nevertheless range anywhere from no carpet to 100% carpet. In effect, each simulated person is assigned to a simulated house with specific properties that affect the potential for exposure.

Although each input variable is assigned once per person, the actual amounts of time spent in each area are variable in SHEDS. Suppose f_carpet = 0.75 for a given person. This means that on average this person spends 75% of their in-home time in carpeted areas. But on any one diary event, the time on carpet might be nothing, or the whole diary event, or some portion of the diary event. This is randomly sampled and is determined using the "ramp function" discussed in the next section.

The f_AreaTreated variable is used to divide the in-home time between treated and untreated portions of the home. If the user has no contrary information, the times may be equated to fractional areas. For example, if one knows that a given scenario typically results in 25% of the home being treated, then it is reasonable to assume that a person will spend (on average) 25% of their in-home time in this area. However, if there is reason to suspect that the time fractions will differ from the area fractions, then the former should be specified. SHEDS defines the f_AreaTreated variables as fractions of the house rather than as actual areas for several reasons. First, if area were sampled directly then sometimes the treated area would exceed the house size. When sampling the fraction, the natural bounds are zero and one. Second, larger houses probably have larger treated areas, so the fraction may be more universal (that is, it applies to more houses) than any actual area. Finally, SHEDS uses the variable as a time fraction, so entering it as a fraction provides a clearer interpretation; that is, fractions near one mean that nearly all in-home time is spent in treated areas, while fractions near zero indicate otherwise.

2.6.2.1 Partial treatment of lawns and gardens

If a user wishes to model partially treated lawns and gardens, then there are two ways to modify the SHEDS input data. Suppose a fraction "f" of the lawn is treated. Then the user may reduce the time fraction input variable f_lawn by this factor, or may calculate an average concentration over the entire lawn, which would be a fraction "f" of the standard initial concentration. Either way, the average exposure from this scenario would be reduced by a factor of "f". However, if would not be correct to decrease both the concentration and the time fraction by this factor, as the exposure is proportional the product of these two terms and hence would be reduced by a factor of "f" squared. Similar remarks apply to partial garden treatments, except that the relevant input is the garden time fraction f_garden.

2.6.3 The “ramp function”

The purpose of the “ramp function” defined in SHEDS is to transform a uniform U(0,1) random number into a time fraction. The ramp function always has the same basic appearance: it consists of three regions. In the first region (small values of the random number), it returns a time fraction of zero. In the second or intermediate region, the returned value increases linearly from zero to one (this constitutes the "ramp"). In the third region (high values of the random number) the ramp function returns a time fraction of one.

Besides the U(0,1) random number, each call to the ramp function requires two non-random inputs. The first is the mean value of the returned time fraction, which is one of the variables from Table 2-6. This mean may range from zero to one, inclusive. Over a sufficiently large number of calls, the mean output from the ramp function will approach this value. The second non-random input determines the size of the intermediate or ramp region: a value of zero means the ramp becomes a step function (only the first and third regions remain), while a value of one means that the intermediate region is a large as possible, given the first requirement (the target for the mean output value).

In SHEDS, the ramp function is defined as

$$\text{ramp}(u, f, r) = \min(\max((u - 1 + f + r*f) / (2*r*f), 0), 1), \quad \text{if } f \le 0.5$$
$$= \min(\max((u - 1 + r + f - r*f) / (2*r*(1-f)), 0), 1), \quad \text{if } f > 0.5$$

Here "u" is the uniform random number, "f" is the mean value of the output, and the third parameter "r" is a setting which controls the frequency of "mixed" output (that is, returned values that are neither 0 nor 1, but something in between). The default setting in SHEDS is r=0.5. When r=0 or f=0 the above forms involve division by zero so the following alternatives are used:

$$\text{ramp}(u,0,r) = 0$$

$$\text{ramp}(u,1,r) = 1$$

$$\text{ramp}(u,f,0) = 0, \text{ if } u < 1-f$$
$$= 1, \text{ if } u \ge 1-f$$

The ramp function is used in SHEDS to assigns event duration fractions. For example, if the person is in-home on a diary event, some of that time will be in carpeted area and some on hard floors. The average fraction of in-home time in carpeted areas is a SHEDS input distribution, and a specific value is assigned to each person. But there would be insufficient time variation in behavior if the person divided every diary event into the same fractions of carpet time and hard-floor time. The ramp function is used to assign the carpet time fraction to each in-home diary event. Note that if the user sets the ramp size parameter "r" to zero, then an in-home diary event is either all on hard floors (if the ramp returns 0), or all on carpet (if it returns 1), although this decision may change each time the random number "u" is resampled. Similarly, the ramp function is used outdoors to apportion lawn and garden time, and for time in the presence of the pet. A test of this parameter was made for quality assurance purposes (test QA15 in Chapter 8), and varying it had a very small effect on the mean exposure.

The "u" variables for the ramp functions are generated automatically by the SHEDS code and are resampled every hour. The "f" values are the constants selected for each person for the input variables listed in Table 2-6. The "r" value affects the variance in the output, and must be the same for all persons and all calls to ramp function in a given model run. For greatest consistency with SHEDS version 3, "r" should be close to ½, so this is the default value in version 4. This follows because the mean diary event duration in the CHAD database (for non-sleeping events) is close to 30 minutes, which meant that in version 3 a random yes/no test would be independently sampled twice per hour (that is, once on each diary event). If the chance of "yes" is 50%, the expected outcomes would be 25% both results are "no," 25% both are "yes," and a 50% chance of one "no" and one "yes." Using the ramp function with f=0.5 and r=0.5, there is a 25% chance of returning a time fraction of zero, a 25% chance of returning a fraction of one, and a 50% chance of splitting the time among the two outcomes.

2.6.4 Example using a generic indoor scenario

Suppose for a given person the input variables have the following values:

f_carpet = 0.6
f_AreaTreated = 0.3

If these values are sampled from distributions, then potentially each person may be assigned different values. The f_carpet value is the same for all indoor scenarios, but f_AreaTreated may differ for each one.

Suppose that for a given event the random u values are 0.5 for the carpet test and 0.7 for the treated test. The ramp functions then return values of 0.75 and 0.5, respectively. This means that a fraction 0.75 (or 75%) of the duration of this in-home event is spent in carpeted areas, while the remaining 25% is automatically assigned to hard floors. The ramp function for the "treated" time returns a value of 0.5, which is the value of ramp(0.7,0.3,0.5), which means that half of the event is is a treated area and half in an untreated area. In effect, each in-home diary event has its duration split into four mutually exclusive categories: treated carpet, untreated carpet, treated hard floors, and untreated hard floors. For other in-home diary events, the u values supplied to the ramp functions will generally be different, so the resulting time splits will also be different.

For a single diary event, all the other variables in the transfer equations (except for concentrations) are fixed, whether the person is in a treated area or an untreated area. However, the transfer rates may differ on carpets and hard floors, so those concentrations cannot be combined until the SHEDS transfer equations are used (in the exposure module). Thus, a weighted average of the treated and untreated concentrations may be taken. For example, using the transfer coefficient (TC) approach, the transfer equation on carpet would be

$$\text{Transfer} = TC_{carpet} * C_{treated\_carpet} * T_{treated\_carpet} + TC_{carpet} * C_{untreated\_carpet} * T_{untreated\_carpet}$$

$$= TC_{carpet} * T_{event} * (C_{treated\_carpet} * Tf_{treated} * Tf_{carpet} + C_{untreated\_carpet} * Tf_{untreated} * Tf_{carpet})$$

Here the C variables are surface concentrations, the TC are transfer coefficients, the T variables are durations, and the Tf variables represent the time fractions. T_{event} is simply the diary event duration. $Tf_{treated}$ is the return value of the ramp function which used f_AreaTreated as an input, while $Tf_{untreated} = 1 - Tf_{treated,}$ since all indoor time must be either in a treated or in an untreated area (so the fractions must sum to one). Tf_{carpet} is the return value from the ramp function call that used f_carpet as an input. Given the ramp function values, the term in parentheses in the above equation may be calculated directly in the SHEDS concentration module. That weighted average is the only concentration variable needed in this transfer equation. In the example given above, the weighted average concentration is 75% of the unweighted average of the treated and untreated carpet concentrations, reflecting the fact that 75% of the event duration is spent on carpet.

Furthermore, if there are multiple indoor scenarios, then the term in parentheses above may be summed over all scenarios in the concentration module. It would be inappropriate to sum all the scenario-specific $C_{treated\_carpet}$ values directly, since the person may encounter mostly treated carpet under one scenario but mostly untreated carpet under another scenario. However, once these weighted averages have been taken, then such sums become valid. If the TE (transfer efficiency) approach is used for transfer, then the same argument applies.

Similarly to the carpet example shown above, a weighted average concentration for hard floors may be obtained. The transfer equation is

$$\text{Transfer} = TC_{hard} * C_{treated\_hard} * T_{treated\_hard} + TC_{hard} * C_{untreated\_hard} * T_{untreated\_hard}$$

$$= TC_{hard} * T_{event} * (C_{treated\_hard} * Tf_{treated} * Tf_{hard} + C_{untreated\_hard} * Tf_{untreated} * Tf_{hard})$$

Of course, $Tf_{hard} = 1 - Tf_{carpet,}$ since the two floors types cover all cases. The term in parentheses may be evaluated directly in the concentration module, and summed over all indoor scenarios. In the above example, the weighted average concentration is 25% of the unweighted average of the two hard-floor concentrations, since 25% of this particular diary event is spent on hard floors.

In the SHEDS exposure module, the final transfer equation for dermal contact with the floor is

$$\text{Transfer} = TC_{carpet} * T_{event} * C_{carpet} + TC_{hard} * T_{event} * C_{hard}$$

Two points should be noted here. First, it does not matter whether the transfer coefficients (or transfer efficiencies) are the same on the two surfaces or not; the same equation applies in any case. Second, only the total event duration (not any time fractions) appears in the above equation. Since this is always recorded on the diaries, the implementation of the transfer equations in the exposure module becomes very simple. The complications arising from variable contact durations have been moved to the concentration module, as have any issues relating to multiple scenarios. No loop over scenarios is needed in the exposure module.

Just like the carpet and hard floors, a weighted average of air concentration is obtained using the time fractions in the treated and untreated areas:

$$C_{air} = C_{treated\_air} * Tf_{treated} + C_{untreated\_air} * Tf_{untreated}$$

This concentration C_{air} applies on activity diary events in the in-home micro (which means all events in the home except those in which the person is sleeping).

2.6.5 Example lawn or garden scenarios

For diary events in the out-home micro, time may be spent on the lawn, in the garden, or neither. SHEDS assumes that the entire lawn and/or garden are treated (if they are treated at all). Hence, the average fraction of out-home time in treated areas is given by

f_AreaTreated = f_lawn + f_garden

It is not logical for this sum to exceed one, since the assumption is that one cannot be on the lawn and in the garden at the same time. If the sample values for these variables do exceed one for some person, then they are proportionally reduced to prevent this. The ramp function is then called using this value of f_AreaTreated and the appropriate randomly generated u-variable. The event-specific return value from this ramp function gives the time fraction for the given out-home event that is spent in a treated area, either lawn or garden. A second ramp function is called which splits this time between the lawn and garden. This is analogous to the indoor situation, where a ramp function call splits the time between carpet and hard floors.

For example, suppose that the values assigned to this person are:

f_lawn = 0.6
f_garden = 0.2

This means that an average of 80% of out-home time will be spent either on the lawn or in the garden (since 0.6+0.2=0.8), and that 75% of this "treated" time is spent on the lawn (since 0.6/0.8 = 0.75). For each diary event, two calls are made to the ramp functions to determine the actual time fractions that are assigned. This method of combining the lawn and garden time fractions is preferred to sampling them independently, since it prevents the sum of the two from exceeding one.

The transfer for this out-home diary event becomes

$$\text{Transfer} = TC_{lawn} * C_{lawn\_scenario} * T_{lawn} + TC_{garden} * C_{garden\_scenario} * T_{garden}$$

$$= TC_{lawn} * T_{event} * (C_{lawn\_scenario} * Tf_{treated} * Tf_{lawn}) +$$

$$TC_{garden} * T_{event} * (C_{garden\_scenario} * Tf_{treated} * Tf_{garden})$$

Here $Tf_{garden} = 1 - Tf_{lawn,}$ which prevents the person from being in both places at the same time. The Tf variables are the values returned by the ramp function calls. The terms in parentheses are the duration-weighted concentration averages for the lawn and garden, respectively. These can then be summed over all the lawn and garden scenarios, to obtain the averages C_{lawn} and C_{garden}, respectively.

In the exposure module, for any out-home events the dermal transfer becomes

$$\text{Transfer} = TC_{lawn} * T_{event} * C_{lawn} + TC_{garden} * T_{event} * C_{garden}$$

As noted for the indoor scenarios, this transfer equation (which appears in the exposure module) uses only the total event duration, and does not refer to any specific scenarios. Again, it does not matter whether the TC values are the same or different for the lawn and for the garden.

2.6.6 Example pet scenario

The pet scenarios differ from the others in that they can occur simultaneously with indoor or outdoor scenarios. For example, it is possible to contact the pet at the same time as the carpet (or hard floor), whereas the times on carpet and on hard floor are mutually exclusive.

For all events in the in-home or out-home micros (which excludes sleeping events), a random portion of that time is considered to be "near the pet." This is determined by a single call to the ramp function using f_pet as the mean output value. The return value becomes the time fraction used in the duration term in the transfer equation. If transfer coefficients are used, then the transfer from the pet is

$$\text{Transfer} = TC_{pet} * C_{pet\_scenario} * T_{pet}$$

$$= TC_{pet} * T_{event} * (C_{pet\_scenario} * Tf_{pet})$$

The Tf value is the value returned by the ramp function. If there are multiple pet scenarios, then the duration-averaged concentrations (the terms in parentheses above) are summed over the pet scenarios to obtain an overall time series C_{pet}.

In the exposure module, for all events in the in-home or out-home micros, the transfer from pets is given by

$$\text{Transfer} = TC_{pet} * T_{event} * C_{pet}$$

As for the other types of scenarios, this transfer equation only uses the full event duration (not the actual contact time), and is not scenario-specific.

2.6.7 Summary of duration-weighted concentration averaging

For all types of scenarios, contact duration is factored into the duration-weighted average concentrations, which are obtained using the scenario-specific concentrations and the results from the ramp function calls. A total of seven hourly concentration time series are produced (for each person) by the concentration module: C_{carpet}, C_{hard}, C_{lawn}, C_{garden}, C_{pet}, C_{air}, and C_{sleep}. If the activity diary places the simulated person in their home, then the person is dermally exposed to the sum of ($C_{carpet} + C_{hard} + C_{pet}$), and receives inhalation exposure to C_{air}. In the out-home micro, the person is exposed to the sum of ($C_{lawn} + C_{garden} + C_{pet}$). When sleeping, the person is exposed only to $C_{sleep.}$ Elsewhere, the simulated person receives no new exposure.

The concentration module in SHEDS initially produces hourly time series for each combination of scenario and medium. These are reduced in number by averaging, using the time fractions produced by the calls to the ramp function. If there are five indoor scenarios, then the ramp function is called nine times for each simulation hour, per person. The returned values (for example, the fraction of out-home time spent on the lawn) may change hourly. This fits nicely with the fact that the initial concentration time series are evaluated hourly, as are the seven averaged time series that are passed to the SHEDS Exposure Module.

Once the appropriate weights have been included for each scenario, SHEDS sums the concentration time series over scenarios, restricted to the chemical currently being processed in the model. Some scenarios might supply nothing because they do not contain the current chemical, or they might not affect the given medium. This summation produces an overall hourly concentration time series for each of the following: indoor air, carpet, hard floor, lawn, garden, pets, and sleep. The "sleep" micro is simply the untreated air concentration, summed over indoor scenarios. Even if the bedroom floor has chemical on it, it is assumed that the bedding does not, so there is no dermal exposure during sleep. A total of just seven hourly concentration time series are needed for calculating exposure, regardless of the number of scenarios selected.

2.6.8 Changes from version 3 in the determination of contact

SHEDS version 3 had a list of 18 input variables to determine the media contacted, most of which were conditional probabilities such as "probability of being in a treated area while at home and awake". There were two difficulties with this; first, the number of such inputs was rather large, and second, some users found the variables rather obscure and could not identify appropriate data sources.

In version 4, each indoor scenario may have its own treated area. Some scenarios may treat the entire house, but others may be applied to a small area. When multiple scenarios are modeled in the same

version 4 run, variables such as "probability of being in a treated area while at home and awake" have to become scenario-specific, which increases the number of inputs beyond the 18 required in version 3. Hence, a method for simplifying the specification of contact probabilities was developed for version 4.

One change in version 4 was to reduce the number of areas in the house from three to two. In version 3 there was a treated area where the chemical was directly applied, an untreated area to which the chemical could disperse, and a third area with zero chemical concentration (too far from the treated area to receive appreciable dispersion). In version 4 this third area is eliminated. The user can still create areas with zero concentration in the modeled house by assigning zero dispersion to the untreated area. Otherwise, the untreated area in version 4 should be considered to be the average of all the non-treated parts of the house. Obviously, in a real house these will not all have the same concentrations anyway, as areas adjacent to a treated area will receive more chemical than those farther away. So in both version 3 and version 4, the concentration of the "untreated area" represents the average of several areas. In version 4, this average includes the areas with zero concentration.

Another change in version 4 is that the chemical in dust or soil is no longer distinguished from chemical in other forms, in terms of the potential for contact. In version 3, on a single indoor diary event it was possible to contact dust only, residue only, or both, or neither. In version 4 the chemical in dust is effectively mixed with the chemical in other forms to produce an overall chemical loading on the floor. Then if contact occurs, part of this total loading is transferred to the person. However, this requires fewer contact-related input variables since the dust and residue contacts do not have to be determined separately.

A third change in version 4 is that the remaining input variables relating to contact are expressed in forms that are more easily visualized. For example, the variable F_Carpet is simply described as "the fraction of the house that is carpeted", but a better description would be "the fraction of in-home time spent in carpeted areas, while awake". In the absence of information to the contrary, one would assume that these two quantities are the same. But if they are not, then the latter is the quantity which is needed in SHEDS. For many users, the latter quantity will be unknown, but the former may be obtained from various sources, or at least an "expert opinion" may be elicited. Both version 3 and version 4 effectively require the latter definition of this variable, but version 4 may express it as the "fraction of the house that is carpeted" to make it more accessible to the user. Similarly, the scenario-specific "fraction of house that is treated" really refers to the time fraction spent in the treated area, but may be reinterpreted as an area fraction, if that makes the data acquisition easier.

2.7 Generating Exposure and Dose Time Series

2.7.1 SHEDS equations for exposure by pathway

The estimation of exposure is the primary function of the SHEDS model. Exposure is defined in SHEDS-Residential as the contact between a chemical agent and a human target at the skin surface or oral/nasal boundary. The skin exposure surface is separated in SHEDS into hands and the rest of the body. Dose is the amount of chemical that crosses these boundaries. The time series for exposure and dose preserve within-day peaks and variability as a person moves through time (Figure 2-5).

In SHEDS, air exposures are equal to the chemical concentration in air and are measured in micrograms per cubic meter (µg/m^3, or ug/m3 in code or printouts). The inhalation dose is the mass of chemical that is inhaled, measured in micrograms per diary event. Daily totals of the inhaled dose are also calculated (with units of ug/day or mg/kg/day). SHEDS variables using units of mg/kg/day are indicated by the prefix "DD" (for "daily dose") in their names.

There are two measures of dermal exposure in SHEDS. The “new exposure” is the mass of chemical transferred from the contact media to the person on the current diary event (equivalent to the “exposure mass” definition in Zartarian et al., 2005). New exposure only occurs during contact events with media that have non-zero concentration. The other measure of exposure is called the “running exposure,” which is the exposure mass currently on a given exposure surface. In SHEDS, once contact between the agent and target at the exposure surface occurs, the chemical is assumed to be retained on the exposure surface until it is absorbed, washed off, transferred to another body part, or otherwise eliminated. In this manner, the effects of increased washing are manifested in SHEDS by decreases in time-averaged running exposure and decreases in absorption and dose. However, washing has no effect on the amount of new exposure for a given diary; it merely reduces the running exposure.

New dermal exposure is measured in micrograms per diary event. To obtain a daily total for new exposure, SHEDS simply sums it over the diary events. Running dermal exposure is expressed in micrograms, but it should not be summed over events since the "same" chemical mass may be counted on multiple events. Instead, a duration-weighted average is appropriate, to obtain the average running exposure over a day. Dermal dose is the chemical mass penetrating into the *stratum corneum*, the outermost layer of the skin, measured in micrograms per diary event. Note that not all of the SHEDS dermal dose will penetrate all the way through the skin, which may affect comparisons with other models or measurements.

There are two forms of non-dietary ingestion in SHEDS: hand-to-mouth transfer and object-to-mouth transfer. Both use the same rules for exposure and dose. Exposure is the mass of chemical that is placed inside the mouth. When there are multiple insertions, the masses are summed. For example,

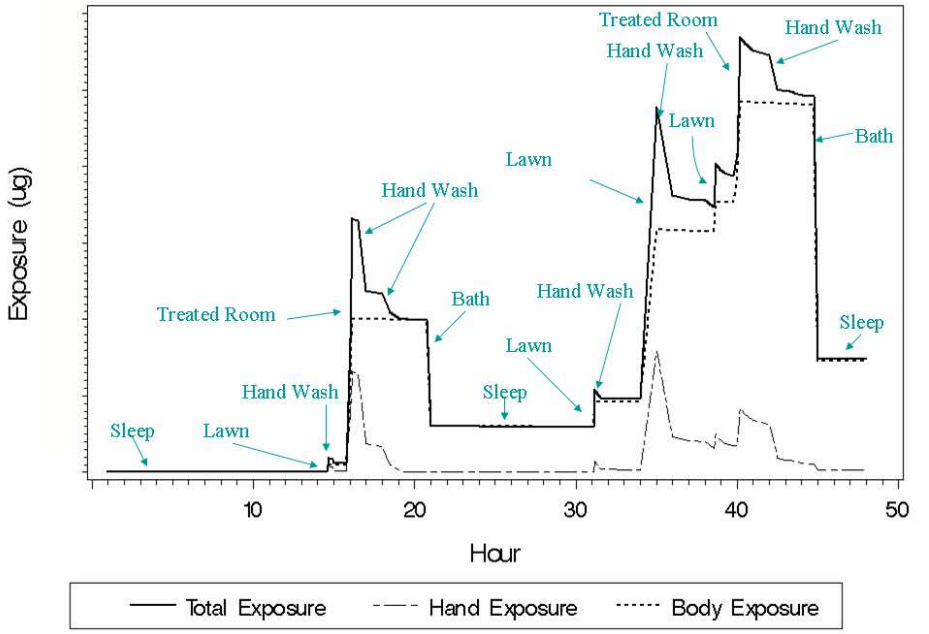


Figure 2-5. Example time profile of dermal exposure from surface residues.

repeatedly sucking a thumb may result in more "exposure" than the mass of chemical on the thumb. The dose is the amount of chemical removed from the hand or object, which then enters the gastrointestinal (GI) tract, measured in micrograms per diary event.

Due to differences in units, it is difficult to summarize exposure across pathways. (See Table 2-7.) But the doses are all measured in micrograms chemical crossing the exposure surface per diary event, for the inhalation, dermal, hand-to-mouth, and object-to-mouth pathways. SHEDS calculates a total dose which is the sum of these components.

Table 2-7. Residential exposure pathways in SHEDS

Symbol	Description
Inh	Inhalation of air
Hands	Chemical on hands
Body	Chemical on rest of skin
Der	Hands+Body
Htm	Hand-to-mouth transfer
Otm	Object-to-mouth transfer
Ing	HTM + OTM

2.7.1 Inhalation Exposure and Dose

Equation 2-2

$$E_{inh,\,e} = C_{a,e}$$

$$D_{inh,\,e} = C_{a,e} * BVR_{e} * METS_{e} * T_{e}$$

where

$E_{inh,\,e}$ = inhalation exposure for a given diary event *e* [µg/m^3]
$C_{a,e}$ = airborne concentration of chemical for the diary location [µg / m^3 air]
$D_{inh,\,e}$ = inhalation dose for a given diary event *e* [µg/event]
BVR_{e} = basal ventilation rate for the individual (accounts for body weight) [m^3 air / hour]
$METS_{e}$ = activity-specific ventilation rate ratio [-] (multiple of basal rate)
T_{e} = duration of diary event [hour/event]

The inhalation exposure is simply the average air concentration during the event. The inhalation dose is the amount of chemical crossing (entering) the oral/nasal boundary, which is the product of the air concentration, the breathing ventilation rate, and the event duration. The breathing ventilation rate is a product of two factors: the basal rate is constant for a given person, whereas the METS value is an activity-dependent multiplier of the basal rate.

The METS values are sampled from activity-specific distributions, listed on the "mets" input file. This file can be edited in SAS, but not directly through the SHEDS interface. The default METS distributions are the same ones used in the CHAD database. Note that the “raw” METS values are

adjusted according to the rules given by Isaacs et al. (2008). These adjustments account for two main effects: the reduction in METS that becomes necessary when people get tired (due to long sequences of high METS values), and the effective increase in METS due to excess post-exercise oxygen consumption (EPOC), which corresponds to heavy breathing or "panting" after exercise ceases. Another adjustment is made for differences in health, as measured by mets_max, the person-specific maximum allowed mets value. Basically, the underlying mets distributions are compressed for those with low mets_max, and are extended to higher values for those persons with high mets_max.

2.7.2 Dermal exposure and dose

2.7.2.1 New Dermal Exposure

The SHEDS user has two options for entering the variables used to calculate dermal exposure: the transfer efficiency (TE) method and the transfer coefficient (TC) method. A given model run must choose one of these methods exclusively. Both are ways to determine the amount of chemical moving from the environment onto the skin surface.

The model internally uses TC to estimate exposure. Depending on the user's choice, TC can be entered directly, or the model will calculate TC by multiplying a dermal transfer efficiency (TE) by the skin surface area contacted per time (cm^2/hr) (Zartarian et al., 2005).

Dermal transfer coefficients, with units of area per time, are typically derived empirically by obtaining the mass collected on a cotton dosimeter (e.g., gloves, bodysuit) worn by an individual contacting chemical for a known time (e.g., while picking apples, doing exercise in a treated room, conducting typical activities in a residence or day care center) (µg/hr) and dividing that by an average residue contacted (µg/cm^2).

Dermal transfer efficiency represents the fraction of chemical on a contacted surface area that is transferred onto the skin. If the user selects the TE approach, then SHEDS internally calculates an equivalent TC, using the skin-surface area and contact rate:

Equation 2-3

$$\mathbf{TC_{eff,\,b,e} = A_e * CR_{b,e} * TE_e}$$

where

$TC_{eff,\,b,e}$ = effective transfer coefficient for body part *b* diary event *e* [cm2/hr]
A_e = surface area of skin [cm^2]
$CR_{b,\,e}$ = skin-to-media surface contact rate for body part *b*[1/hr]
TE_e = transfer efficiency [-]

Here $CR_{b,e}$ represents the fraction of the dermal area contacting the surface per hour. This value may exceed unity, implying multiple contacts for the same skin area. For example, a CR value of 2.0 means that over a one hour period, the entire uncovered skin area averages two contacts with surfaces. The body surface area reflects the amount of skin that is not covered by clothes. The fraction of the body uncovered is sampled once per day from an input distribution. It is assumed that the hands are always uncovered. The hands and body share the same value for the transfer efficiency, although they will generally have different CR values. If the transfer coefficient method is selected by the SHEDS user, then a single transfer coefficient is used for the total skin surface, with a fraction F_DermalHands going to the hands. Thus, when using transfer coefficients,

$TC_{eff,hands,\,e}$ = TC_e* F_DermalHands

$TC_{eff,body,\,e}$ = TC_e* (1-F_DermalHands)

Different surface collection devices are available (e.g., aluminum plates, rollers, sleds, hand presses, hand wipes, surface wipes); some collect dislodgeable material and some collect the total applied chemical. When developing inputs for either TE or TC, it is important that the choice reflect the measure of surface loading (that is, dislodgeable or total).

Regardless of whether the transfer efficiency or transfer coefficient approach is used, equation 2-4 determines the new dermal exposure, or amount of chemical transferred onto the skin:

Equation 2-4

$$\mathbf{E_{b,e} = C_{surf,e} * TC_{eff,b,e} * T_e * f\_skin_{b,e} * f\_load_{b,e}}$$

where

$E_{b,e}$ = new dermal exposure for body part *b* on diary event *e* [µg / event]
$C_{surf,e}$ = available loading of chemical on the surface contacted on diary event *e* [µg / cm^2]
$TC_{eff,b,e}$ = effective dermal transfer coefficient on diary event *e* [cm^2 / hr]
T_e = duration of diary event *e* [hr / event]
$f\_skin_{b,e}$ = adjustment for clothing on body part *b* on diary event *e* [-]
$f\_load_{b,e}$ = adjustment for pre-existing dermal loading on body part *b* on diary event *e*[-]

The adjustment f_skin reflects the effects of clothing. SHEDS assumes that the hands are always uncovered, so f_skin_{hands} always equals one. For the body this fraction is f_uncloth, the fraction of the body area that is not covered with clothing. For assessing unit handler exposures it is assumed that the handler is wearing shorts and short sleeves, which corresponds to f_uncloth near one-half. Hence, the unit handler exposures are multiplied by (f_uncloth/2), to account for changes in exposure potential due to the amount of clothing.

The adjustment f_load reflects a reduction in net transfer due to pre-existing dermal loading. In version 3, dermal loading proceeded at its "base" rate until the maximum dermal loading was reached, whereupon any additional loading was assumed to immediately be lost. This is modified in version 4 to have decreasing net transfer as the dermal loading is increased. In version 4 the maximum dermal loading still exists, and if it is exceeded then the excess is immediately lost, as in version 3. Cohen Hubal et al. (2005) and Nishioka et al. (2003) support the concept of decreasing transfer with multiple successive contacts. The loading adjustment is:

Equation 2-5

$$\mathbf{f\_load_{b,e} = max\ (1-RunExp_{b,e}\ /(Dermal\_maxload*SA_b),\ 0)}$$

where

$f\_load_{b,e}$ = transfer adjustment factor for body part *b* on diary event *e* [-]

$RunExp_{b,e}$ = running exposure on body part *b* on diary event *e* [ug]

Dermal_maxload = maximum dermal loading limit [ug/cm2]

SA_b = surface area for body part *b* [cm2]

The max function prevents f_load from becoming negative, which otherwise would happen when the running dermal exposure RunExp_dermal exceeded the limit. When this situation exists, new exposure is not allowed, and the excess is removed from the skin. This removal is recorded by the variables Remove_H_maxload and Remove_B_maxload for the hands and body, respectively.

2.7.2.2 Running Dermal Exposure

Running exposure is carried over from one diary event to the next. New exposure adds to running exposure, while washing, bathing, dermal absorption, hand-to-mouth transfer, and brush-off may lower running exposure. Running exposure serves two purposes in SHEDS: it is the basis for the amount of hand-to-mouth transfer and for the amount of dermal absorption. It would be possible to assign the fate of new dermal exposure immediately, but it would then be difficult to assess the effect of, say, subsequent hand washing on either the amount transferred to the mouth or on the amount dermally absorbed. SHEDS considers competing removal processes, with each one applying to the current running exposure. A change in the rate of one removal process automatically is reflected in the amounts subject to the other processes.

The running exposure on diary event *e* is given by

Equation 2-6

$$\mathbf{RunExp}_{b,e} = \mathbf{RunExp}_{b,e-1} + \mathbf{E}_{b,e} - \mathbf{MaxL}_{b,e} - \mathbf{Abs}_{b,e} - \mathbf{HTM}_{e} - \mathbf{Brush}_{b,e} - \mathbf{Bath}_{b,e} - \mathbf{Wash}_{,e}$$

where

$RunExp_{b,e}$ = running exposure on body part *b* on diary event *e* [ug]

$RunExp_{b,e-1}$ = running exposure on body part *b* on diary event (*e*-1) [ug]

$E_{b,e}$ = new exposure for body part *b* on diary event *e* [ug]

$MaxL_{b,e}$ = reduction on diary event *e* for being over máximum loading limit [ug]

$Abs_{b,e}$ = absorption (binding) in *stratum corneum* on diary event *e* [ug]

HTM_{e} = hand-to-mouth transfer of chemical on diary event *e* [ug]

$Brush_{b,e}$ = brush-off of loading on body part b on diary event *e* [ug]

$Bath_{b,e}$ = loading removal from body part b on diary event *e* due to bath/shower [ug]

$Wash_{,e}$ = hand loading removal on diary event *e* by hand washing [ug]

Each adjustment to the loading is calculated once per diary event, in the order shown above. Each diary event starts with the final dermal loading from the prior event. The first adjustment is the addition of new exposure from contact with surfaces. Next is the reduction to the loading limit, if the new exposure caused the loading to exceed its limit. Third is the absorption or binding of the chemical into the *stratum corneum*, the outermost layer of the skin. Next is the reduction due to hand-to-mouth activity. Fifth is the brush-off term, representing mechanical removal of chemical from the skin, e.g., due to contaminated skin contacting a clean surface (this can be set to zero). Finally, the bathing and hand-washing adjustments are made. Bathing is checked first, and if it does not apply then hand-washing may

occur. Figure 2-6 depicts these adjustments to dermal exposure. The amount of chemical entering the *stratum corneum* is the dermal dose and is output to a file to be used as input by an external PBPK model.

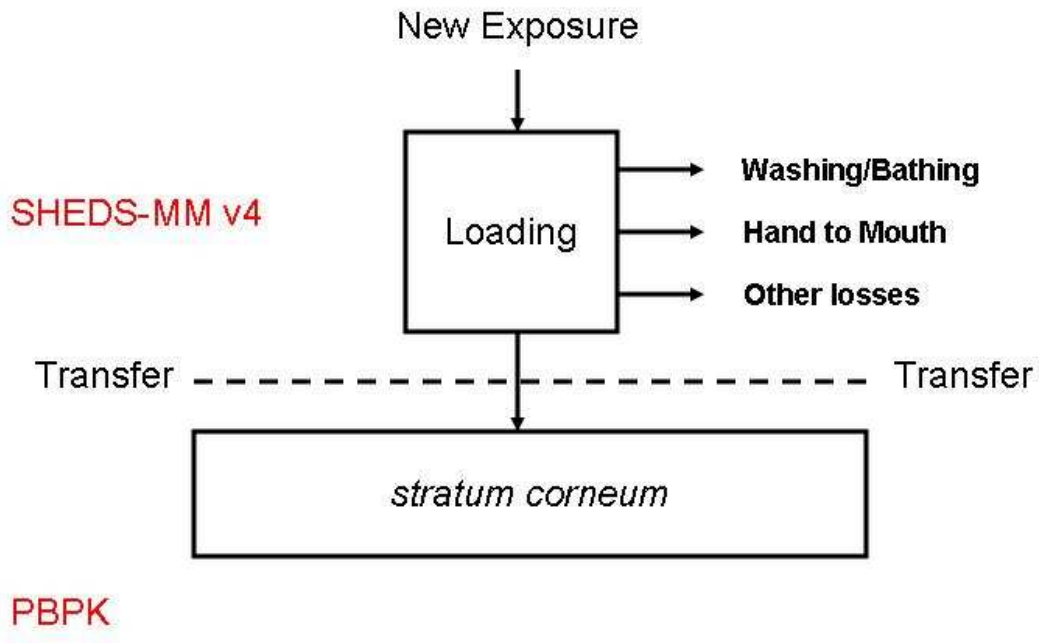


Figure 2-6. Illustration of running dermal exposure in SHEDS-Residential

2.7.2.3 Maximum dermal loading limit

If the addition of new exposure puts the dermal loading over its upper limit, then the loading is immediately reduced back to the limit. A separate check is made for the hands and the body. The adjustments are recorded as the SAS variables Remove_H_Maxload and Remove_B_maxload, for hands and body, respectively. The adjustments are

Equation 2-7

$$\mathbf{MaxL_{b,e} = Max(RunExp_{b,e} - Maxload * SA_b, 0)}$$

where

$MaxL_{b,e}$ = Adjustment for excess loading on body part *b* on diary event *e* [ug]

$RunExp_{b,e}$ = Running exposure on body part *b* on diary event *e* [ug]

Maxload = Maximum dermal loading limit [ug/cm2]

SA_b = Skin surface area for body part *b* [cm2]

2.7.2.4 Dermal absorption

In version 4, dermal absorption refers to the crossing of the chemical from the external skin surface into the *stratum corneum*, the outermost layer of the skin. As a modeling assumption that once the chemical enters the *stratum corneum*, it is no longer subject to external removal processes (e.g., hand mouthing, skin washing). For this reason in SHEDS, dermal absorption and dermal dose are equivalent. Once this occurs, the affected chemical is subtracted from the running dermal exposure (which measures chemical outside the skin surface). This definition differs from version 3, in which the dermal absorption referred to the amount entering the blood after passing through the skin.

On each diary event, the dermal dose is the amount absorbed into the skin, and is given by

Equation 2-8

$$\mathbf{Abs_{b,e} = D_{der,b,e} = RunExp_{b,e} * Binding * T_e}$$

where

$Abs_{b,e}$ = absorption in body part *b* on diary event *e* [ug/event]

$D_{der,b,e}$ = dermal dose for body part *b* on diary event *e* [ug/event]

$RunExp_{b,e}$ = dermal loading on body part *b* on diary event *e* [ug]

Binding = fraction of dermal loading binding to *stratum corneum* [1/hr]

T_e = duration of diary event *e* [hr]

A check is made that the product of the last two factors does not exceed one. In practice, the dermal_binding is usually much smaller than one, so this is seldom an issue.

2.7.2.5 Hand-to-mouth exposure transfer

Equation 2-9

$$\mathbf{HTM_e = D_{htm,e} = (RunExp_{hands,e} / 2) * H\_frac_e * [1 - (1-H\_TE)^{(H\_freq_e * T_e)}]}$$

$$\mathbf{E_{htm,e} = D_{htm,e} / H\_TE}$$

where

HTM_e = hand-to-mouth transfer of chemical on diary event *e* [ug]
$D_{htm,e}$ = new hand-to-mouth ingestion dose for diary event *e* [µg]
$E_{htm,e}$ = new hand-to-mouth ingestion exposure for diary event *e* [µg]
$RunExp_{hands,e}$ = running exposure (on both hands) for diary event *e* [µg]
H_Frac_e = fraction of one hand that enters the mouth on diary event *e* [-]
H_TE = mouthing removal efficiency (fraction transferred to mouth) [-]
H_freq_e = frequency of hand-mouth activity on diary event *e* [mouthing events/hr]
T_e = duration of diary event *e* [hr]

The adjustments to the dermal loading due to hand-to-mouth transfer are explained below in section 2.7.3. The first assumption behind the hand-to-mouth transfer is that each diary event starts with the dermal hand loading spread evenly over the two hands. The division by 2 gives the chemical mass on one hand. Each diary event is likely to have several hand-mouth contacts. Within a diary event, it is assumed that repeated hand mouthings all involve placing the same part of the hand into the mouth, and that no reloading occurs. Each repetition therefore involves less mass remaining on the hand and

therefore less transfer. For example, three thumb insertions in one diary event does not result in triple the mass transfer of a single insertion, because the thumb has been partially cleaned prior to the second and third.

The term in square brackets in Equation 2-9 gives the fraction of the dermal mass entering the mouth that is removed (that is, transferred to the GI tract). For a diary event of duration T_e hours, the expected number of hand mouthings during this event is ($H\_freq_e \times T_e$). Instead of determining the mass removed, consider the mass that remains on the hand. This remaining mass is multiplied by (1- H_TE) every time mouthing occurs. As an example, suppose H_TE = 20%, or (1- H_TE) = 0.80. Then after three mouthings, a fraction 0.80^3= 0.512 of the chemical mass that was mouthed remains on the hand. After N mouthings, the fraction left behind would be (1- H_TE)^N, and for this diary event N = ($H\_freq_e \times T_e$). The fraction transferred to the mouth is one minus the fraction left behind, giving the factor in Equation 2-9.

The term ($H\_freq_e \times T_e$) may not be an integer. Suppose it turns out to be 3.7 on some diary event. An exponent of 3.7 will remove more than an exponent of 3, but will remove less than an exponent of 4. The 'fractional' number of mouthings can also be interpreted as 3 full mouthings (each of H_frac_e) and one partial mouthing (an area somewhat less than H_frac_e).

Note that, at most, only an amount ($RunExp_{hands,\ e}$ /2 * H_frac_e) can be removed by hand-to-mouth transfer during one diary event. For example, if H_frac_e = 20% then a maximum of 10% of the hand loading can be removed. However, at the end of each diary event, the remaining loading is redistributed so more could be removed on the next diary event.

The transfer equation assumes that each mouthing event removes the same percentage of the amount of chemical that enters the mouth. The net transfer decreases with repeated mouthings because the hand loading decreases. This is automatically adjusted at the end of each diary event (essentially, the dermal loading is redistributed uniformly). The hand-to-mouth exposure is the total amount of chemical entering the mouth. If the "same" chemical enters multiple times (because it was not removed on the previous mouthings), then it is counted again. For example, 10 hand mouthings with an average of 20 ug of chemical on the part of the hand entering the mouth would constitute an exposure of 200 ug, whether or not it was the "same" 20 ug each time.

2.7.2.6 Dermal brush-off

The dermal brush-off term removes dermal loading on every diary event, representing incomplete carryover of loading from one event to the next. If the rate is set to zero (e.g., because of limitations in available data) then no brush-off occurs. Otherwise, the rate is specified as a fraction removed per hour. The same brush-off rate applies to the hands and body, although the dermal loadings and the amount removed will usually differ. The brush-off equation is:

Equation 2-10

$$\mathbf{Brush_{b,e} = RunExp_{b,\ e} * BR_e * T_e}$$

where

$Brush_{b,e}$ = loading removed from body part *b* on diary event *e* [ug]
$RunExp_{b,\ e}$ = dermal loading on body part *b* on diary event *e* [ug]
BR_e = brush-off rate [1/hr]
T_e = duration of diary event *e* [hr]

2.7.2.7 Bathing and hand washing

The running dermal exposure in SHEDS may be reduced by bathing and/or hand washing. "Bathing" is a general term that includes showers and swimming, and removes a fraction of the total running exposure. Hand washing removes a fraction from the hands only and does not affect the rest of the body.

Bathing events are sometimes indicated as such on the original CHAD activity diary. CHAD was assembled from a dozen or more separate studies, and these were not consistent in how bathing events were recorded. Some studies simply included these among "personal time" or some similar non-specific description. A longitudinal SHEDS diary will generally be constructed from multiple original studies, and the consequence is that it may have some baths, but there may also be long stretches without them. To counter this irregular pattern, each simulated individual in SHEDS is assigned a value for the maximum number of days between baths. Once this time is reached without an explicit bathing event, a bath is forced onto the end of the current diary event. The fraction of loading removed by bathing is given by the input variable "f_RemovBath." This applies to both the hand loading and body loading.

Hand washing tends to be more frequent than bathing, but only applies to the running exposure on the hands. Hand washing is never explicitly indicated in CHAD and must be added by the model. First, each person is assigned a random value for the "HandWash_mean" variable, drawn from the user-specified distribution. In SHEDS, hand washing may take place only once in each clock hour, so HandWash_mean actually indicates the number of hours in which hand washing occurs. Two or more hand washes in the same hour count as one for modeling purposes. HandWash_mean may have a non-integer value, as it represents an average over many days.

SHEDS assumes that the probability of hand washing during any (waking) hour is HandWash_mean/16. At the start of each hour, an independent random check is made. If the check succeeds, then a minute is selected at random from within the current hour. If the person is awake at that time (and not bathing), then hand washing is added to that diary event. If they are asleep, the check fails (which explains why the denominator in the hourly probability is 16 rather than 24). The actual number of hand washes will vary from day to day, as will their timing.

Bathing has a higher priority than hand washing in SHEDS calculations. SHEDS implicitly includes hand washing as part of bathing and does not allow both adjustments on the same diary event. Both bathing and hand-washing remove a simple fraction of the current running exposure. For bathing this is

Equation 2-11

$$\mathbf{Bath_{b,e\_} = RunExp_{b,e} * f\_RemovBath}$$

where

$Bath_{b,e\_}$ = loading removed from body part *b* on diary event *e* [ug]

$RunExp_{b,e}$ = dermal loading on body part *b* on diary event *e* [ug]

f_RemovBath = fraction of loading removed during bath/shower event [-]

For hand washing the amount removed is

Equation 2-12

$$\mathbf{Wash_{e\_} = RunExp_{hands,e} * f\_RemovHandWash}$$

where

Wash $_{e\_}$ = loading removed from hands on diary event *e* [ug]

RunExp $_{hands,\ e}$ = dermal loading on hands on diary event *e* [ug]

f_RemovHandWash = fraction of hand loading removed during hand washing [-]

2.7.3 Non-Dietary Ingestion pathway

There are two sources of non-dietary ingestion in SHEDS. One is hand-to-mouth (HTM) transfer, which is described above in the adjustments to the dermal loading. The second is object-to-mouth (OTM) transfer, detailed below. This represents the behavior of picking up toys and other objects that may have been lying on the floor and mouthing or chewing them. The chemical concentration on the surfaces of such objects is assumed to be proportional to the concentration on the floor (this ratio is the variable F_OF) If toys were not present when the room was treated, F_OF should be much less than one. Conversely, if the objects were present when the room was treated, one might expect F_OF=1. The user should decide which is more appropriate for their modeling scenario.

The exposure-dose boundaries have been moved in version 4 from those used in version 3. Now, the exposure surface is the mouth, and any chemical that contacts the mouth is an exposure. Chemical that is transferred to the mouth (and then to the GI tract) is called ingestion dose. If there are multiple hand-to-mouth (HTM) contacts, then there are multiple HTM exposures and the exposure is summed. In this manner, the HTM exposure may exceed the amount of chemical on the hands.

The determination of the HTM exposure on each diary event is complicated because the amount of chemical left on the hands diminishes with each mouthing event. However, the HTM exposure and HTM dose still follow a simple relationship, because on each HTM contact the same fraction of chemical is transferred to the mouth. If this fraction is x%, then the HTM dose on any diary event will be x% of the HTM exposure on this same event. Hence, by calculating either one, the other may easily be obtained.

The same rules apply to object-to-mouth (OTM) exposure and dose. Multiple OTM contacts might result in an OTM exposure that exceeds the amount of chemical present. However, the OTM dose will never exceed the amount of chemical present, and the OTM dose will always be x% of the OTM exposure, if the OTM transfer efficiency is x%.

The formulas in SHEDS version 4 for the HTM and OTM doses are the same as for the exposure transfer (the chemical mass moving from the hand to the GI tract) in version 3. Instead of modifying these formulas to calculate HTM and OTM exposure, the latter are back-calculated from the HTM and OTM doses by dividing by the appropriate transfer efficiency.

Object-to-mouth exposure transfer

Equation 2-13

$$\mathbf{D_{otm,\ e} = C_{surf,e} * f\_OF * O\_area_e * [1 - (1-O\_TE)^{(O\_freq_e * T_e)}]}$$

$$\mathbf{E_{otm,\ e} = D_{otm,\ e} / O\_TE}$$

where

$D_{otm,\ e}$ = object-to-mouth ingestion dose for event *e* [µg]
$E_{otm,\ e}$ = object-to-mouth ingestion exposure for event *e* [µg]

$C_{surf,\ e}$ = chemical concentration on contacted surface for event *e* [µg/cm^2],
F_OF = object-to-surface concentration ratio [-]
$O\_area_{\ e}$ = area of the part of the object that enters the mouth on event *e* [cm^2]
O_TE = fraction transferred to mouth from object surface per mouthing event [-]
$O\_freq_{\ e}$ = frequency of object-to-mouth contact on event *e* [mouthing events/hr]
$T_{\ e}$ = duration of diary event *e* [hr]

The object-to-mouth transfer uses the same logic as the hand-to-mouth transfer presented above. These objects are generally assumed to be toys that children leave on the floor and other exposed places, later to be played with and chewed. The amount of chemical on the toy depends on the variable "F_OF" (called F_ObjFloor in the SAS code), which gives the typical concentration ratio for toy surfaces to the floor. If toys or other objects are not targeted for treatment, but acquire chemical through contact with the floor, this ratio should be less than one.

The exposure and dose for the non-dietary ingestion route are the sums of the HTM and OTM contributions, from equations 2-9 and 2-13 respectively:

$$E_{ing,\ e} = E_{htm,\ e} + E_{otm,\ e}$$

$$D_{ing,\ e} = D_{htm,\ e} + D_{otm,\ e}$$

As noted above, SHEDS provides exposure outputs which may be used with detailed PBPK models. Figure 2-7 illustrates SHEDS ingestion exposures (contact at the dashed line representing the mouth) and doses (represented by arrows crossing the mouth into the GI tract), both non-dietary and dietary.

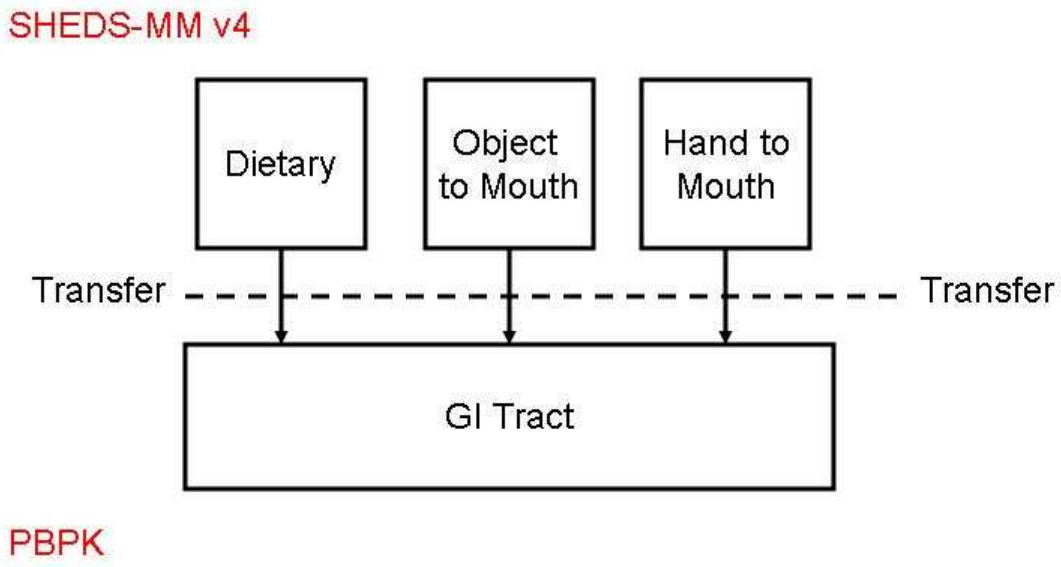


Figure 2-7. Illustration of dose in SHEDS-Residential

2.8 Handler exposures

A new feature in version 4 is the estimation of exposure during handling or application of the scenarios. Two quantities are estimated: the amount of new dermal exposure (mass of chemical added to the skin loading) and the inhaled dose (the amount of chemical breathing in through the nose or mouth). Each of these is determined using a "unit exposure factor" which represents the ratio of mass affecting the individual to the mass applied to the environment. The units are micrograms on the individual per gram in the environment.

The chemical mass in the environment is the product of an "application rate" and the area treated. Generally, scenario-specific guidance is given on the product label for the application rate, and larger treatment areas require more chemical to be used. For indoor scenarios the treated area is expressed as a fraction of the house area (that is, the floor area). Some scenarios may cover the entire house, but others are applied to limited areas. For outdoor and pet scenarios, the treated area is assumed to be the given size of the target location. The basic handler equations are:

Equation 2-14

$$H_{hands,p} = U_{der,p} * AppRate_p * AreaTreated_p * F\_Hands * F\_skin_{hands} * F\_load_{hands}$$

$$H_{body,p} = U_{der,p} * AppRate_p * AreaTreated_p * (1-F\_Hands) * F\_skin_{body} * F\_load_{body}$$

$$H_{inh,p} = U_{inh,p} * AppRate_p * AreaTreated_p$$

where

$H_{hands,p}$ = new handler exposure on hands when applying scenario p [µg / usage]
$H_{body,p}$ = new handler exposure on body when applying scenario p [µg / usage]
$H_{inh,p}$ = handler inhalation dose when applying scenario p [µg / usage]
$U_{der,p}$ = unit dermal exposure factor for handlers of scenario p [ug/g]
$U_{der,p}$ = unit inhalation exposure factor for handlers of scenario p [ug/g]
F_Hands = fraction of total dermal exposure during handling that goes to hands [-]
F_skin_b = adjustment factor for unclothed skin for body part *b* [-]
F_load_b = adjustment for running exposure for body part *b* [-]
$AppRate_p$ = chemical application rate for scenario p [g/m2]
$AreaTreated_p$ = area treated with scenario p [m2]

The dermal unit handler exposure assumes no protective clothing and no pre-existing dermal loading. To be consistent with the estimation of post-application exposures in SHEDS, the two adjustment factors f_skin and f_load that were used in Equation 2-4 are also applied. No adjustments are made to the inhalation exposure (this assumes no dust mask or similar device is worn).

2.9 Daily exposures

The time series of event-level exposures and doses are output to files intended for input to a PBPK model. These files are quite large, and it is not practical to maintain all the SHEDS output data in them. Once the files with saved information have been written, SHEDS produces daily totals for the exposure and dose variables. All subsequent calculations are performed on the daily totals.

There are two ways to form daily values. One is to simply sum the event-level values across all the events on each day. The other way is to take duration-weighted averages across events. For example, new dermal exposure is best measured with a daily sum over events, which measures the total amount of chemical transferred from the environment to the skin each day. If the diary events had been broken into

shorter intervals, the new exposure on each one would be less (because the formula includes time duration), but the total would not necessarily change. On the other hand, the running dermal exposure is better represented by a daily average.

Most exposure and dose variables are represented by daily sums. There are three exceptions: running dermal exposure, inhalation exposure, and breathing ventilation rate. These quantities represent rates or averages, even within a diary event, and should as well on a daily basis. Other variables subject to duration-weighted averaging include chemical masses and concentrations in various media, and variables that express rates, fractions, or efficiencies.

2.10 Blood and Urine doses

As mentioned earlier, SHEDS outputs diary event-level time series of the doses (the amounts of chemical crossing the dermal, ingestion, and inhalation exposure boundaries). These files can be used as input to a detailed PBPK model. SHEDS does not attempt to properly track the chemical once it has entered the body, but it does include a simple, built-in calculation to provide the user with estimates of blood and urine doses. Note, however, that the focus of SHEDS is the exposure module; a more sophisticated PBPK model may be needed for the user's intended purpose.

The SHEDS blood dose is a weighted sum of the three route-specific daily doses. The user supplies the weights, which are mass ratios. For example, if 70% of the chemical that binds to the skin eventually enters the blood, then the weight should be f_DermalBlood = 0.7. For the lungs, f_InhalBlood represents the ratio of the chemical mass inhaled (entering the nose or mouth) to the mass absorbed into the blood (almost entirely in the lung alveoli). This will be less than 100% because some of the inhaled air enters "dead space," and additionally, some chemical will remain in exhaled air, even if it entered the lungs. For ingestion, this ratio would exclude any chemical that is not bioavailable (meaning that it is in a non-digestible form), or is otherwise not absorbed.

The SHEDS blood dose represents the mass of chemical entering the blood on a daily basis. The formula is

Equation 2-15

$$\mathbf{D_{blood,d} = D_{der,d} * f\_DermalBlood + D_{ing,d} * f\_IngestBlood + D_{inh,d} * f\_InhalBlood}$$

The second subscript "d" on each dose variable reflects a daily total. The mass of chemical entering the on a daily basis is

Equation 2-16

$$\mathbf{D_{urine,d} = D_{blood,d} * f\_Urine}$$

Like other doses in SHEDS, $D_{blood,d}$ and $D_{urine,d}$ are measured in micrograms (in this case, per day).

.

2.11 Daily Dose variables and Margin of Exposure (MOE)

Daily dose, or "DD variables" in SHEDS version 4 refer to daily dose variables that are in units of (mg/kg/day), the units most commonly used in dose and risk estimation. Other SHEDS dose variables report chemical mass in micrograms. For example, the relationship for dermal dose is:

Equation 2-17

$$\mathbf{DD_{der,d} = D_{der,d} / (1000*weight)}$$

where

$DD_{der,d}$ = daily dermal dose for day d (mg/kg/day)
$D_{der,d}$ = SHEDS dermal dose for day d (ug/day)
1000 = units conversion from micrograms to milligrams
weight = body weight for this person (kg)

It should be noted that each simulated individual has a specific (and generally unique) body weight. Hence, the simulated individual with the highest dose $D_{der,d}$ might not be the one with the highest $DD_{der,d}$ and vice versa. Furthermore, while the above formula applies similarly to the inhaled and ingested doses (and the sum of the doses), it is not correct to apply it to statistics of the dose distribution. For example, the 95th percentile of $D_{der,d}$ is not related to the 95th percentile of $DD_{der,d}$ by the above formula.

In addition to daily dose variables, there is one "daily exposure" or DE variable, for the new dermal exposure. It is calculated similarly to dermal dose, namely:

Equation 2-18

$$\mathbf{DE_{der,d} = E_{der,d} / (1000*weight)}$$

where

$DE_{der,d}$ = daily dermal exposure for day d (mg/kg/day)
$E_{der,d}$ = SHEDS dermal new exposure for day d (ug/day)
1000 = units conversion from micrograms to milligrams
weight = body weight for this person (kg)

The units for DE_{der} (DE_dermal in the code) are also (mg/kg/day). This variable is calculated in SHEDS because it may be needed for the MOE calculations (see below).

Average daily dose (ADD) is another commonly used statistic in dose modeling. This is also reported in (mg/kg/day), and can be obtained from the DD variables by a straightforward average (within each person) over the days in the simulation period.

A margin of exposure (MOE) is the ratio of a reference exposure (or dose) to an actual exposure (or dose). There is one MOE for each main route (dermal, inhalation, and ingestion), and another for the total across these routes. The dermal MOE is generally taken to be a ratio of dermal exposures rather than loadings, although both forms are known. If the dermal MOE is based on the dermally absorbed dose, then the numerator is generally taken to be the same as for the ingestion and inhalation routes. When the dermal MOE is based on external measures such as exposure, then the numerator may be substantially larger than for the other routes, since dermal absorption is relatively small.

The numerator in the MOE calculations is called the "NOAEL," or No Observed Adverse Effect Limit. See US EPA (2007) for further discussion of NOAELs. Whether the numerator is a NOAEL or another calculated reference dose, SHEDS requires them to be provided in units of (mg/kg/day). SHEDS allows the user to input a separate NOAEL for each of the three routes. The MOEs are

Equation 2-19

$$\mathbf{MOE_{der} = NOAEL_{der} / DE_{der}}$$

$$\mathbf{MOE_{inh} = NOAEL_{inh} / DD_{inh}}$$

$$\mathbf{MOE_{ing} = NOAEL_{ing} / DD_{ing}}$$

SHEDS also calculates the reciprocals of each MOE variable (reported as the InvMOE variables), for use in the total MOE equation below. The total MOE is found by reciprocal addition:

Equation 2-20

$$MOE_{tot} = 1 / (1/MOE_{der} + 1/MOE_{inh} + 1/MOE_{ing})$$

This calculation is performed for each person-day.

2.12 Output data sets

Most of the SHEDS output data sets contain the chemical name as part of their file name. For illustration of example file names, the chemical name will be assumed to be X. The output data sets are written to the \output directory under the run name and chemical name. For example, if the user submits a job called "test1," then a \test1 directory is created in the location specified by the variable "RunPath" in the RunInfo file. The general input for this run is placed in \test1\input, while the chemical-specific input is placed in \test1\X\input. For a multi-chemical run, there may be multiple directories under the same run name (for example, \test1\Y\input will contain chemical-specific input for chemical Y in the same run. The output for chemical X is written to \test1\X\output, while the output for chemical Y is written to \test1\Y\output.

Both the event-level and the daily summary file are saved for the last person in the run. The event-level file (for chemical X) is called "expo_X" and the daily summary file is "daily_X." If the user specifies person numbers, then the event-level and daily files for each of these persons are saved, using the person number as a suffix on the respective file names.

The "for_PBPK_X" file can be very large. It contains only seven variables (person number, day number, event duration, vent_inhal, dose_inhal, dose_ingest, and RunExp_dermal). However, there is one record for every event, for every person. For 1000 persons and 1 year each, this amounts to 15 million records, occupying almost 1 gigabyte. A large run may be 10,000 or more persons. Setting the switch "for_PBPK" to zero on the RunInfo file prevents this output file from being created. This is most likely necessary on sensitivity or uncertainty runs.

The "alldays_X" file contains one record for every person-day. It may be produced in either of two modes. The smaller version (produced with the option Save Intermediate Variables = 0) contains 49 variables per record, while the larger version (with Save Intermediate Variables = 1) contains 112 variables. The larger version occupies about 300 megabytes for a run of 1000 persons.

There are five other files written to the output directory. The "DiaryCounts" file reports the number of activity diaries available for selection in each pool. Unlike all the other output files, this file does not expand as more persons are run. The "Persons_Demog" file contains one record for each simulated person, reporting properties such as age, gender, weight, and so on. These properties were all assigned prior to performing exposure or dose calculations. This file is useful in conjunction with larger output files such as "for_PBPK_X" or "alldays_X" which do not contain constant personal information such as age and gender for space considerations. In a multichemical run, the variables on "Persons_demog" should be the same for all chemicals.

The "Persons_Mean_X" file contains the personal mean daily values for each of the exposure and dose variables. Note that each variable is analyzed separately, so the date on which the mean dermal exposure occurred may differ from the date on which the mean inhalation exposure occurred, and similarly for all the other variables. This means that there was no single "mean day" on which all the

variables were simultaneously at their means. The "Persons_Max_X" file contains the personal maximum daily values for each exposure and dose variable, over the simulation period. As for the personal means, these maxima likely occurred on different days and there would be no simulation day when all (or even most) of the variables were simultaneously at their respective maxima.

The "Loops_Mean_X" file contains means across all person-days on a given pass through the outer loop of a two-stage model run. This is useful in certain types of sensitivity analysis, particularly the percentile scaling approach, which is discussed in a later chapter.

2.13 Population Variability Estimates of Exposure or Dose

The steps described above are applied to each simulated individual. To obtain population estimates SHEDS-Residential repeats this process many times using 1-stage Monte Carlo simulation (Cullen and Frey, 1999) for variability results (see Figure 1-2). Statistical weights derived from the United States Census (U.S. Census Bureau, 2000) are applied so that population sampling is proportional by age and gender to reflect the U.S. population.

SHEDS may be applied to either the entire population, or a selected subpopulation. This subpopulation may be a particular set of age-gender cohorts, or particular chemical usage groups, or both. The age-gender groups are selected by the user at the start of a model run. To model certain groups of "users" the modeler may have to modify the usage probabilities appropriately, and (if necessary) the ‘p_lawn’, ‘p_garden’, or ‘p_pet’ probabilities as well.

The population cumulative distribution functions (CDFs) reflect variability of exposures due to differences in both the time individuals spend contacting chemicals and exposure factors that affect how much of the chemical reaches a child’s body after contact. In addition to producing CDFs and summary statistics tables, SHEDS-Residential computes the contribution to exposure from each of the exposure pathways. This can take the form of variability CDFs for single pathways, or as pie charts that illustrate the relative importance of each pathway for a given population statistic, such as the mean individual.

As with any stochastic process, the variability estimates improve with larger sample sizes (more simulated persons). A sample size of 1000 persons is sufficient to obtain fairly accurate estimates for the center of the variability distribution. For estimating the exposure for extreme percentiles of the population, larger sample sizes are needed in order to generate a sufficient number of examples to locate the percentile points with confidence.

SHEDS can also examine the variability of exposure over time within each simulated person. This is generally driven by chemical usage patterns, with the obvious result that exposures are high for a few weeks following each application, and low at other times. Since results for a single person are dependent on the random draws selected for that person, few clear inferences can be drawn from the exposures time series for a single person. In this context, distributions across many persons are more useful. But when each person is assigned different chemical usage patterns and application dates, the variation in usage will dominate over other effects.

To obtain a clear picture of time variability, it is best to assign fixed application dates to all persons in the model run. Then the model will generate exposure time series for many persons that can easily be combined. For example, one can obtain the population distribution of exposure on an application day, or one day after application, or two days after, and so on.

The same discussion applies to dose variables. The intake or absorbed doses are evaluated at the diary event level and are saved on the "for_PBPK_X" output file. Daily summaries of these variables are on the "alldays_X" file. The built-in PK model in SHEDS produces daily time series of the blood and

urine dose for each of the three main routes, for each simulated person. To examine population variability, one would first evaluate a dose metric of interest for each simulated person, to reduce the time series to a single value. The "Persons_mean" and "Persons_max" files contain two such statistics, but others could be constructed. Then variability distributions (CDFs) of this metric across the selected population can be constructed. One could also construct pie charts showing the contribution of each route to the total dose, for some metric such as the population mean.

Alternatively, one can examine the time variability of dose if one runs the model with fixed application dates (the same for all the simulated persons). Then one could plot distributions of, for example, the distribution of urinary concentration at various numbers of days after an application. The options for dose estimates are similar to those available for exposure estimates. These are discussed further in the section "Population Variability Distributions."

2.14 Coding Topics

This section presents a short overview of SAS code for both the SHEDS residential model and its graphical user interface.

2.14.1 Residential Code

All of the code for running the SHEDS-Residential model (without the GUI) is contained in a single file called Multimedia4.sas. While this code contains several modules, as listed in Appendix C, the code was placed in a single file because the modules are designed to be used together, rather than be ported as is to other programs. Each module has a header which lists the macros it contains.

The SHEDS code contains very little data. Virtually all the hard-coded data in SHEDS refers to physiological variables, such as the relationship between energy expenditure and breathing ventilation rate, or the fraction of the dermal surface area found on the hands. Nearly all data is supplied on input files, including data on the body weight distribution for each age-gender cohort. These input files are all SAS data sets, which may be edited by anyone who has SAS. The more commonly modified input files are explicitly examined and modified in the GUI.

The residential code in version 4 is meant to be more readable than earlier versions. This entails less use of the SAS macro language, and a closer correspondence between the equations in this Technical Manual and in the code. A few differences remain, mainly due to notational restrictions in SAS. For example, no subscripts or superscripts may appear in SAS code, nor any Greek letters.

The SAS language itself is composed of modules, and end users do not have to license all the SAS modules. The SHEDS code requires only the BASE, STAT, GRAPH, and ACCESS modules. A model run could be completed without the GRAPH module, but certain plotting features for the output would no longer be available. The ACCESS module (for PC files) is required only for the import or export of SAS data sets in other formats (for example, as Excel files). SHEDS version 4 was written and tested using SAS version 9.1. It is expected that SHEDS should run correctly using either SAS 8.2 or SAS 9.2, but these have not been subject to quality assurance tests. Therefore, the user should use SAS 9.1 or else may encounter issues if a different SAS version is used.

Each macro in the residential code has an extensive comment section at the end, immediately before the %mend statement. These comments are numbered, corresponding to numbered references appearing throughout the main body of the macro. There are also some single-line comments which appear in the code without numbers.

The SHEDS code makes use of both SAS macro variables and data set variables. SAS permits only one SAS data set to be open at a time, except when data sets are being merged. SAS also permits only one record of a data set to be processed at a time. If information from separate records or separate data sets needs to be referenced together, the most common approach is to store certain values in SAS macro variables. These are basically text strings (possibly containing numbers) that are placed into the code before running each DATA or PROC step. One DATA or PROC step is compiled at a time, after this macro variable substitution has occurred. For example, all the formulas for sampling variables from distributions are stored in SAS macro variables, which can be pasted into the code wherever these variables are referenced. However, while this is very flexible, it makes the source code hard to read, especially when the macro variables themselves contain references to other macro variables (essentially, nesting of macro variables). Such nesting was used fairly extensively in SHEDS version 3, but has been almost eliminated in version 4.

The SHEDS code uses the symbol 'ug' for micrograms, since the Greek letter µ cannot be used. The code classifies the randomly sampled variables into three categories: Demographic, Simulation, and Product variables. Demographic variables can only be sampled once per person. They include obvious variables such as age and gender, but the category includes others that are not really demographic in nature, such as the size of one's lawn. The categories are important because they determine how and when the variables are assigned values in the code. The simulation variables are ones that are expected to vary throughout the simulation period. A file is constructed using the hourly calendar for the simulation period, and each simulation variable is assigned a value for each hour. If a simulation variable varies more slowly (for example, if it is sampled daily), then the same value will be repeated each hour until the day changes. The Product variables are ones that are associated with the use of particular scenarios. While these could be called Scenario variables instead, the category is sometimes indicated internally in SHEDS by the first letter, and Scenario would not then be distinguished from Simulation.

Currently, there are no plans for a SHEDS coder's or programmer's manual. However, the code is organized in a top-down manner, with a short main program that shows the sequence of steps in SHEDS, with clearly indicated nested looping structure over chemicals, uncertainty or sensitivity, and variability. Other than these loops, the SHEDS code is basically linear, with each macro (apart from Utility macros) called from only one place in the code. It would not be difficult to flatten it out by eliminating all macro calls (other than the Utility macros) and placing all the code in the main program. This is not done only because it would obscure the overall structure of the code. In effect, the main program serves as a one-page synopsis of the SHEDS model.

The SHEDS code is placed in the file \prg\Multimedia4.sas during installation. This file is readable (and editable) in any text editor. The SAS editor has a parser that color-codes various types of statements, making the code easier to read. While the user may examine the code without SAS, the model can be run only if the user has SAS installed. SAS is also required to examine or edit the input and output data sets, with the exception of the text output files intended as input to a PBPK model.

2.14.2 The SHEDS-Residential User Interface Code

The SHEDS-Residential interface was created in SAS using the SAS/AF application development platform. The interface consists of forty-seven individual application frames (screens) and their corresponding SAS Screen Control Language (SCL) files. The SCL code in each of these files control the behavior of the frames and how they interact with SHEDS-Residential datasets and libraries. In addition to the frame SCL files, the interface also contains several other SCL code files that are used to perform various tasks, but are not associated with one particular interface frame.

The interface frames and SCL files are packaged in the MM4 SAS catalog, stored in the ./prg directory under the SHEDS-Residential home directory, which is assigned the SAS library **prg** during the execution of the program.

In addition to the interface frames and SCL files, the interface also makes use of custom SAS components (widgets) that have been designed especially for SHEDS-Residential. An example of one of these custom components is the widget for entering probability distributions. This component, which is used many places in the interface, contains code for automatically validating the user's entries and saving the information in a form matching that used in the SHEDS-Residential input datasets. These custom components are packaged in the Components SAS catalog, which is also stored in the ./prg drectory.

When SHEDS-Residential is installed, a shortcut is placed on the user's desktop. Clicking on the shortcut will call the SHEDS-Residential autoexec.sas program (in the home directory), which sets a number of SAS system variables and initiates the interface. While the interface is running, users familiar with SAS will be able to perform many standard SAS tasks, including opening, viewing, and editing SAS datasets and SAS programs (including the SHEDS-Residential code). However, there are certain exceptions and caveats. Note that all SHEDS input datasets that are open in the table viewer should be closed before exiting any interface screen that edits them. If this is not done, there is a chance that the information entered in the interface will not be written to the dataset. Another exception is that the user cannot submit any SAS program or code during the execution of the interface. The user will receive the message "Error: you cannot submit your program until autoexec processing has completed." Exiting the interface (via the main interface screen) will complete the autoexec processing and allow the user to submit programs.

3 MODEL INPUTS

SHEDS-Residential is a very flexible model which may be applied to a wide range of chemical exposure scenarios. To simulate the specific scenario(s) desired, the user must provide a set of appropriate inputs (Table 3-1). These describe, for example, the chemical usage and properties, human behavioral characteristics, and various exposure and dose factors. The job settings section requires either specific numbers or character text strings to be supplied. The other inputs generally allow random sampling from probability distributions, although point values can also be used to fix certain inputs. The abbreviation "Pr" stands for "probability" in Table 3-1, and may refer to either a yes/no (binomial) variable or to a probability vector.

It is important to note that while the total number of SHEDS input variables is large, not all inputs are needed for every SHEDS run. Some are specific to a chosen modeling algorithm (such as Decay or Intervals), while others apply to certain locations (such as outdoors), and may not apply for a given scenario. The simplest approach is to leave the unnecessary inputs at their default values.

Table 3-1. Summary of SHEDS user-specified input variables (not all apply to all model runs).

- Job settings (not randomly sampled):
  - Run Name
  - Run Path (where the inputs and outputs are stored)
  - Chemical name(s)
  - Metabolite name(s)
  - Number of persons to simulate
  - Age-gender cohorts to be modeled
  - Start date for simulation period
  - End date for simulation period
  - Number of uncertainty iterations
  - Random number seed
  - Method for specifying scenario usage
  - Method for determining concentration after usage
  - Dermal transfer method: TC or TE
  - Diary assembly method: eight diaries or D&A method
  - Handlers: all persons, some persons, or none
  - Minimum age to be a handler
  - Maximum permitted # usages for one scenario
  - Maximum permitted # houses (timeseries input only)
  - Diary key variable (D&A method only)
  - Diary diversity D statistic (D&A method only)
  - Diary autocorrelation A statistic (D&A method only)
  - Fraction of diary events in multiple contact locations
  - Number of days prior to simulation period to model concentrations
  - NOAEL for dermal exposure
  - NOAEL for ingested dose
  - NOAEL for inhalation dose
  - Specific person numbers for reporting full output data
  - Flag for expanded or condensed variable lists on output files

- Flag for debug level (number of messages on log)
- Flag for pre-processing the activity diary database
- Flag for conducting percentile scaling sensitivity run
- Flag for conducting Sobol sensitivity run
- Flag for saving log file as permanent file
- Flag for indicating progress on log
- Flag for correlating random input variables
- Flag for producing detailed output file for input to PBPK model

- Activity-Related or house-related inputs:

Variable	Sampling Frequency
Fraction of in-home time spent on carpet	Person
Fraction of out-home time spent on lawn	Person
Fraction of out-home time spent in vegetable garden	Person
Fraction of home time spent near pet	Person
Max # days between baths	Person
Mean # hand washes per day per person	Person
Daily soil ingestion rate	Person
Pr (having a lawn)	Person
Pr (having a vegetable garden)	Person
Pr (having a dog or cat)	Person
Size of house	Person
Size of lawn	Person
Size of vegetable garden	Person
Size of pet	Person
Pr (re-entry time indoors)	Person
Pr (re-entry time outdoors)	Person
Pr (re-entry time for pets)	Person
Dust loading on hard floors	Person
Dust loading on carpet	Person

- Inputs that are both scenario-specific and chemical-specific:

Variable	Sampling Frequency
Application rate (chemical mass / area treated)	Daily (once per application)
Unit dermal exposure of handlers	Daily (once per application)
Unit inhalation dose of handlers	Daily (once per application)
Initial surface concentration after use: hard floors	Daily (once per application)
Initial surface concentration after use: carpet	Daily (once per application)
Initial surface concentration after use: lawn	Daily (once per application)
Initial surface concentration after use: garden	Daily (once per application)
Initial surface concentration after use: pet	Daily (once per application)
Initial concentration in dust (indoor scenarios)	Daily (once per application)
Initial concentration in soil (lawn or garden scenarios)	Daily (once per application)

Fraction lost per day (decay/dispersion method only)	Person
Ratios for untreated to treated room concentrations (indoor scenarios only)	Person
Concentration at various intervals after usage (intervals method only)	Daily (once per application)

- Scenario-specific inputs (but not chemical specific):

Variable	Sampling Frequency
List of scenarios used in this model run	Not randomly sampled
List of chemicals in each scenario	Not randomly sampled
Usage dates (fixed dates method)	Not randomly sampled
Usage hour (fixed dates method)	Not randomly sampled
Re-entry or ban time (fixed dates method)	Not randomly sampled
Pr (# applications per year) (Model-Determined dates method)	Person
Pr (month of use) (Model-Determined dates method)	Daily (once per application)
Pr (day of week) (Model-Determined dates method)	Daily (once per application)
Pr (hour of use) for indoors, outdoors and pets (model-determined dates)	Person
Usage co-occurrence rules: blackout window	Not randomly sampled
Usage co-occurrence rules: probability multiplier	Not randomly sampled
Usage co-occurrence rules: multiplier window	Not randomly sampled
Usage co-occurrence rules: list of scenarios affecting current selection	Not randomly sampled
Pr (being the handler or applicator)	Person
Fraction of house treated for each indoor scenario	Person

- Chemical-specific inputs (but not scenario-specific)

Variable	Sampling Frequency
Outdoor background concentration	Person
Object- floor concentration ratio	Person
Dermal binding (absorption) rate	Person
Fraction of dermal dose entering blood	Person
Fraction of inhalation dose entering blood	Person
Fraction of ingested dose entering blood	Person
Fraction of blood dose entering urine	Person
Chemical/metabolite mass ratio	Not randomly sampled

- Non-Chemical Specific Exposure Factors

Variable	Sampling Frequency
Dermal transfer coefficient (TC)	Hour
Fraction of TC and Handler exposure going to hands	Hour

Dermal transfer efficiency (TE)	Hour
Hand-surface dermal contact rate	Hour
Body-surface dermal contact rate	Hour
Hand-to-mouth transfer efficiency	Hour
Object-to-mouth transfer efficiency	Hour
Maximum dermal loading limit	Person
Fraction of body unclothed	Day
Fraction of hand surface area that enters mouth	Person
Hand mouthing frequency	Day
Object-to-mouth contact surface area	Person
Object-mouth contact frequency	Day
Dermal brush-off rate	Person
Fraction of dermal loading removed during bath/shower	Person
Fraction of hand loading removed during hand washing	Person

Values for the relevant variables needed for a given scenario must all be specified prior to a SHEDS run. In addition, there are other SHEDS input data listed in Table 3-2 that do not require any user action. Changes can be made to these inputs by judiciously altering or replacing the relevant files, but this is intended for advanced users only and is not supported by the SHEDS interface.

.

Table 3-2. SHEDS inputs not directly selected by the user

- Age-gender population data
- Activity diary database
- Height-weight-body surface area relationships
- Basal metabolic rate
- Breathing ventilation rate (METS) algorithms

3.1 Stochastic Sampling

As a Monte Carlo model, SHEDS relies extensively on random or stochastic sampling. It estimates the population variability by modeling many different people who collectively both represent the target population and who meet the specified distributions for a suite of input variables. The user can control many aspects of the modeling through these input variables, but one needs to be careful to specify all these inputs sensibly. SHEDS version 4 includes a default file for non-chemical specific exposure factors based on available literature data; the user can change these as desired. Note that values in the default file should not be assumed to be appropriate for all actual modeling scenarios, and should be checked by the user.

The "Variables" input file lists all the SHEDS variables that are stochastically sampled. This list has some differences from the list of variables for which the user supplies distributions. First of all, some randomly sampled variables (such as age, gender, and body weight) use data from pre-generated input

files but do not directly accept user-specified distributions. The user could still alter these outcomes by editing the data on the appropriate files. For example, new height-weight-gender-age relationships could replace those on the existing input file. For a few random variables (such as the Mets adjustments), the SAS code would have to be edited to effect such changes.

The second reason why the list of distributions differs from the list of random variables is that in some cases the same random variable is applied to several distributions. This occurs when conditions are set that determine when each distribution is used. For example, there could be multiple distributions for hand mouthing frequency that depend on age or even on location (indoors and outdoors could be different). Only one of these distributions applies at any given time, and they all refer to the same underlying random variable. The other case in which multiple distributions use the same random variable occurs when different variables are expected to be very strongly correlated. Using the same random variable produces this relationship by selecting the same percentile from each distribution. Weaker correlations can be obtained by explicitly defining correlation targets between different random variables.

On the Variables file, the "UserDistrib" column indicates the source of distributions for each random variable. When UserDistrib=0 the model does not accept user-supplied distributions. UserDistrib=1 indicates that the distributions should be specified on the Distributions input file. Finally, UserDistrib=2 indicates that other input files (for example, Decay, or Handlers) supply the distributions.

The other important information on the Variables input file is the "SampFreq" column which shows the rate at which random samples are drawn. Table 3-1 shows version 4 input variables and their default sampling frequencies that can be changed by the user. There are three main types of random variables: Demographic, Simulation, and Product. The sampling frequencies cannot be altered for Demographic and Product variables. However, the Simulation variables may be sampled at Person, Year, Month, Day, or Hour rates. When the model reaches a new value for the variable named as the sampling rate (for example, every midnight for the Day sampling), a new random value is drawn from the appropriate distribution.

Stochastic sampling in the SHEDS version 4 SAS code is implemented somewhat differently than in version 3. In version 4, the random sampling assigns values to "u_variables" which are always uniformly distributed between zero and one, regardless of the purpose of the variable in question. The Variables input file supplies all the information needed to generate all the u_variables; specifically, the name of each u_variable and the frequency with which it is resampled. No information about the simulated individual is needed to generate the u_variables. The Distributions and other input files where distributions are specified are used to apply deterministic transformations to the u_variables. These transformations are rank-preserving. For example, if u_weight=0.45 for some person, then that person is assigned the 45^{th} percentile of the appropriate body weight distribution for their age and gender. The rank-preserving nature of these transformations is important when inputs are to be correlated, as actually the correlation is imposed on the u-variables. The rank-preservation means that the final distributions will exhibit the same Spearman correlation as the corresponding u_variables.

Within a single model run, all of the differences between two persons modeled in SHEDS are attributable to differences in their u_variables. Conversely, if two persons in a SHEDS run were assigned the same u_variables, everything about them would be identical, not only age and gender, but also their scenario usage, activity diaries, chemical contacts, exposures, and doses. This does not hold between different SHEDS runs (or even to persons in different outer loops of an uncertainty run) because the input distributions may differ and thus the transformations of the u_variables would differ as well. However, if all inputs are common to two SHEDS runs, including the random number seed than produces the u_variables, then the results will be exactly the same.

The allowed distributions in SHEDS fall into the following categories: point, continuous, probability vector, empirical, and function. Each of these is discussed in turn below. Also, the use of

"conditions" to assign distributions is discussed. Conditions may be used with any of the categories. On the SHEDS input files, the variable "Form" determines the type of distribution; this is not case sensitive and only the first four characters of the "Form" variable are relevant.

3.1.1 Sampling frequency

Each input variable in SHEDS is assigned a sampling frequency. This indicates how often a new random draw is made for that variable. The selected value remains in effect until a new draw is made. For most inputs, the sampling frequencies are not fixed, but are read as input data. All inputs are sampled anew when a new simulated person is begun, so the minimum sampling frequency is once per person. Other options are to sample once per year, month, day, or hour.

There is a trade-off in variance when the sampling frequency is altered. More frequent sampling leads to more within-person variance, but less between-person variance in any quantity that is averaged over time. The sampling frequency should be based on the properties of each input variable. Variables that reflect properties of the residence itself or habitual behavior patterns might be sampled once per person. Variables that change rapidly over time with little difference in means across persons should be sampled once per day (or more often).

The user interface does not support modifying the sampling frequencies. Such modification can be done by editing the appropriate records of the SAS data set “variables” in the default or model run input directory. However, some variables (for example, age and gender) cannot have their sampling frequencies altered without creating both logical and coding inconsistencies.

3.1.2 Conditional distributions

A general feature of the random variables in SHEDS is that the user may specify more than one distribution per variable. The user supplies "conditions" that indicate when each distribution should be used. Conditions take the form of SAS-compatible logical expressions. For example, if a distribution applies only to children under age 6, then the condition is specified as "age<6". These expressions will typically refer to one or more SAS variables (such as "age") and must be consistent with the variable naming and spelling in the code. Age and gender are the most common variables used in conditions, but others can been used as well.

SHEDS implements distributions in general by constructing code using the template "variable = distribution;" (there are a few exceptions, such as empirical distributions). Conditions are included using a prefacing IF statement, the template being "If (condition) then do; variable=distribution; end;".

In general, the user is responsible for ensuring that the conditions cover all possible cases, or else the variable may remain undefined in certain cases. Also, the conditions should not overlap. Otherwise, multiple IF statements would be executed, with the last one overwriting the results from the others. This may lead to unexpected results and should be avoided.

3.1.3 Point values

Point values are the simplest form of input for random variables. The value of the variable does not depend on the associated u_variable at all (although the u_variable is still generated). For example, if one set f_RemovBath to a point value of 0.9, then every time a bath occurred the dermal running exposure would be reduced by 90%, with no variation. Point values are often set to zero, to "turn off" some effect that is otherwise built into the model.

The u_variable is still generated for point values, for two reasons. First, the programming is simpler if the code does not have to "look ahead" to determine which inputs use point values. Second, some inputs may be assigned point values only conditionally. For example, hand-to-mouth frequency may have a point value of zero only for persons over a certain age, but not for others. The u_variable is generated in advance of the age determination. It is most practical to generate it in all cases, even when it is ultimately not needed.

Note that the user may still request correlation between variables when one or both are assigned point values. As noted earlier, the correlation is applied to the u_variables rather than to the final values. However, for point values the u_variable has no influence on the final value, and a point cannot actually be correlated with anything, so the specification of correlation in this case is both irrelevant and potentially misleading. SHEDS does not prohibit such correlation due to the possibility of conditional assignment of distributions. For example, hand-to-mouth frequency is often set to zero for persons above a cutoff age, yet it is possible to request correlation of this variable with others. This really only has meaning for persons below the cutoff age. At present, SHEDS requires the same correlation between any given pair of variables, independent of age or other conditions, due to the fact that the u_variables are generated and correlated before these conditions are evaluated.

Most variables subject to Yes/No outcomes were classified as point values in version 3. For example, the probability of having a lawn was assigned a point value which indicated the likelihood of a *yes* outcome. In version 4 these distributions are called *Binomial*, which is a subtle distinction. The term *point value* is intended to mean that all persons are assigned the same outcome, and this outcome does not depend on the associated u_variable. Neither of these conclusions applies to the binomial input variables, so they were reclassified and now are more like probability vectors.

3.1.4 Continuous Probability Distributions

Table 3-3 lists the continuous probability distribution types and parameters used in SHEDS. The parameters are given generic names v1, v2, and v3 in SHEDS. Detailed descriptions of all the distribution types are provided in Appendix E.

Table 3-3. SHEDS continuous probability distribution types.

Type	Domain	v1	v2	v3	restrictions
Beta	0<x<1	shape1	shape2	[-]	v1>0, v2>0
Exponential	x>v1	minimum	Mean	[-]	v2>v1
Gamma	x>0	shape	Scale	[-]	v1>0, v2>0
Lognormal	x>0	geometric mean(GM)	geom. std. dev.(GSD)	[-]	v1>0, v2>1
Normal	any x	Mean	Standard deviation	[-]	v2>0
Triangular	v1<x<v3	minimum	Peak	max.	v1≤v2≤v3
Uniform	v1<x<v2	minimum	Maximum	[-]	v1<v2
Weibull	x>0	Shape	Scale	[-]	v1>0, v2>0

All of the above distributions listed above have two parameters except for the triangular. The beta extends from x=0 to x=1, and cannot be shifted to other bounds at present. Note that the GSD for the lognormal must exceed unity. For the triangle, v2 may be equal to either v1 or v3, or in between them, but v1 must strictly be less than v3. In SHEDS, only the first four characters of the distribution type are examined by the code, so it does not matter whether "triangle" or "triangular" is used, for example.

Most SHEDS random variables allow continuously distributed output values and are sampled from the above distributions. Usually these inputs are sampled independently, although it is possible to request correlated inputs, as discussed below.

3.1.4.1 Truncation of distributions

SHEDS distributions may be truncated at either or both ends. Usually truncation is applied to continuous distributions, especially to forms like the normal which are otherwise unbounded. In practice there are nearly always limits beyond which the values make no physical sense, and the model should be prevented from continuing with such values (at least until they are brought within range).

Currently there are three truncation options in SHEDS:

Option	User specifies	Action of SHEDS code
1	Bounding values	Values outside bounds are "piled up" at nearest bound
2	Bounding percentiles	Values outside bounds are "piled up" at nearest bound
3	Bounding percentiles	Values outside bounds are effectively resampled

Option 1 is the same as in version 3, where it was the only available choice. The user may specify a minimum and/or a maximum allowed output value. Any random value sampled from the specified distribution that falls outside these limits is moved to the nearest limit. Note that the input parameters (for example, the mean and standard deviation for the distribution in question) refer to the properties of the *untruncated* distribution. Truncation will definitely lower the variance of a distribution, and may affect the mean as well. Truncation of a large part of a distribution is permitted, but it may be misleading since the parameters would no longer correctly reflect the output.

Option 2 is like option 1, except that the user specifies percentiles instead of actual output values. For example, one might request that 2% of the area of a normal distribution be chopped off each end. To do this, set the minimum percentile to 0.02 and the maximum percentile to 0.98.

Option 3 uses percentiles like option 2, but instead of "piling up" values at the bounding points, the values are effectively resampled until they lie within the allowed range. This does not actually require resampling, because the u_variable is linearly transformed to the specified range. For example, to remove 2% of the area at each end, the transformation is:

u_variable = 0.02+0.96*u_variable

The new u_variable (the left side of the above equation) is now uniformly distributed between 0.02 and 0.98. This generates exactly the same distribution that would be obtained by resampling the target distribution until it lies between the values corresponding to the bounding percentiles. Furthermore, the rank-correlation is unaffected by this transformation (that is, the X^{th} percentile before the transformation remains at the X^{th} percentile after the transformation), so the rank correlation of this input with another input is unaffected by the truncation.

Truncation may be used with non-continuous distributions, although the effects are not as clear. These will be discussed below. Truncation of point values should be avoided because it either has no effect at all or else leads to a contradiction (if the point value itself is out of bounds). The SHEDS code does not check for every possible problem. Some checks are performed by the SHEDS graphical user interface (GUI), but in general the user is responsible for avoiding illogical instructions to the model.

3.1.5 Probability vectors

Certain input variables in SHEDS are restricted to discrete values or outcomes. Probability vectors are lists of probabilities that total to one. The appropriate u_variable is used to select one location from the list. The return value from a probability vector is the selected position.

A variable with just two possible outcomes is called a *binomial* distribution in SHEDS. These are quite commonly used for "yes/no" questions, such as "Does this person have a pet?" For a binomial variable, the form is specified as "Binomial" and the probability of a *yes* outcome is given by the v1 parameter. The v2 and v3 parameters are not required.

When there are more than two possible outcomes, the form is specified as "Probvector" on the input file (although the code only checks for the first four characters, namely "Prob"). Then the v1, v2, and v3 columns are not used, and the list of probabilities is entered as a comma-separated string in the character variable "probvector." In such cases, the code interprets the meaning of the returned value (the position in the list). For example, a Month of Usage vector has length 12, and the result indicates the calendar month (with 1 representing January, and so on). Each scenario will have its own Month of Usage vector if the Variable-dates option is used.

The full list of input variables for which the user must explicitly supply probability vectors is given in Table 3-4. Optionally, the re-entry times may be sampled from probability vectors, but empirical distributions are more practical since the length of the probability vectors might become unwieldy, unless reentry times were restricted to small numbers of hours.

Other inputs could be assigned probability vectors in principle. The output from any probability vector is an integer indicating the selected position in the list. Hence probability vectors should only be used for variables for which such output is sensible. Some possible examples are: hand-to-mouth contact frequency, object-to-mouth contact frequency, and mean number of hand washes per day.

Table 3-4. List of user-supplied probability vectors (not including binomial inputs).

Variable	Purpose
NumApps	Probability of exactly 1, 2, 3, … usages of the same scenario at the same house in one calendar year, given that at least one occurs at this house. In batch mode this vector can have any length, but the interface limits it to 12 per year. There is one NumApps variable for each scenario in the model run.
Month	Fraction of applications (usages)of this scenario occurring in each calendar month. This vector has length 12, with the first cell representing January, the second February, and so on. For example, 0.05 in January means that 5% of all applications of that scenario are made in January, regardless of the length of the simulation period. Note that applications that occur outside the simulation period may result in chemical concentration that persists until the simulation period begins. There is one Month variable for each scenario in the model run.
DayOfWeek	Fraction of applications of this scenario occurring on each day of the week. This vector always has length 7, with the first cell representing Sunday, the second Monday, and so on. The model assumes that the same fractions apply throughout the year. There is one DayofWeek variable for each application type used in the model run.
TimeOfDay(outdoors)	The hour of the day when an outdoor scenario is applied. To keep the length of the vector and the randomly-generated application times reasonable, the hours are restricted to 6 a.m. to 5 p.m. The same vector is used for all outdoor scenarios.
TimeOfDay(indoors)	The same as TimeOfDay(outdoors), but applied to all indoor scenarios.
TimeOfDay(pet)	The same as TimeOfDay(outdoors), but applied to pet scenarios.
BathDays	The maximum number of days between baths. The CHAD diaries are inconsistent with regard to recording baths/showers. If the model assembles a longitudinal diary with a long stretch between baths, then baths are added to diary events whose duration causes the accumulated time since the previous bath to equal or exceed the value of the BathDays variable. This is an integer number of days with a maximum value of 7.

3.1.6 Empirical data

These are new in version 4. Distributions consisting of empirical data are implemented by having the user supply a list of possible values. The program then selects one of these values at random, assuming that each value is equally likely. If the user wants certain values to be selected more often than others, then the same values should be listed multiple times, in proportion to the desired selection frequency.

Any random variable that requests user input may use an empirical data distribution. If the user wants to correlate a variable using empirical data with another variable, then it is necessary that the empirical data set be sorted into increasing order. Otherwise, the values may appear in any order.

Each empirical distribution should be in its own SAS data set. All such data sets are placed in the \emp directory, which is a sub-directory under the main \input directory. Each empirical data set should contain one column (that is, one SAS variable) with the same name as the SHEDS variable in question. The data set may contain as many records as the user wants. SHEDS will count the records and then select one record at random. If there are N records, then the one selected is given by ceil(N*u_variable), where (as always) u_variable is uniformly distributed between zero and one.

The empirical data set may have any name that uniquely identifies it. Often, this is simply the name of the SHEDS variable itself. However, if one wants to use multiple empirical data sets which are associated with conditions for use (see the section on conditions below), then these data sets must be given unique names, even though they all represent the same SHEDS variable. The name of the data (not including the folder name or the SAS data set extension) is supplied as the "empfile" variable on the Distributions input file.

There is essentially no limit on the number of samples in an empirical data set. Certainly, millions of samples (and possibly even billions) can be supported. At some point the SAS “ceil” function may fail to separate integers differing by one, at which point certain records might never be selected. Since SHEDS uses direct access methods to read one record, large data sets are not noticeably slower than small ones.

3.1.7 Functions

This is the most general class of input distributions. The default inputs for SHEDS never use functions, and most users are not expected to need to use them either. Functions operate by allowing the user to supply SAS code as text on the input file. This code may refer to any SAS variables that have values at the point in the SHEDS code where the evaluation takes place. This requires knowledge of the SHEDS programming code, so is intended for advanced users only. It is not supported by the SHEDS interface.

The SAS code that defines the function is written to a macro variable which is later pasted into the appropriate SHEDS module. This function may contain multiple SAS statements (meaning that the function may contain semicolons that are used to separate SAS statements). This is an extension from SHEDS version 3 which allowed the user to supply a formula that could be pasted into a single SAS statement, but could not contain multiple SAS statements.

As a simple example, suppose variable "b" is already defined in the current data step, and "c" is assigned the function (2*b). Then variable c will be assigned twice the value of variable b. In principle,

every input could be written as a function, specifically one that transforms the appropriate u_variable. Conditional distributions may be written as functions that contain multiple IF statements. However, the "function" class allows the option of using modeling variables as well as u_variables, and allows the possibility of using multiple variables in the definition.

3.1.8 Correlation of Inputs

SHEDS Multimedia version 4 permits randomly sampled variables to be generated with correlation. As mentioned in section 2.14.1, the random variables in SHEDS are classified internally into three categories: Demographic, Simulation, and Product. Any variables from the first two categories may be correlated with each other. There is no restriction on sampling frequency for correlating variables. When two variables of differing sampling frequencies are correlated, that means that all the pairs of values that apply together (for the same person at the same time in the simulation period) exhibit the desired correlation.

The model requires Spearman or rank correlations as input. SHEDS actually produces the correlation among the u_variables, which are the only items in SHEDS that are randomly generated. The Spearman correlation is unchanged by the rank-preserving transformations that take the u_variables to the final distributions. However, if the final distribution contains discrete values with lots of ties, the ranks of these tied values are somewhat ambiguous and the sample correlation may be affected.

The major reason for excluding the Product category is that those variables are not associated with specific simulation times when the u_variables are first generated. Since the correlation is induced at that point, it would be difficult to ensure that the correlated values were used at the same point in time within the simulation.

SHEDS uses a modified NORTA approach (Smith and Glen, manuscript in preparation). In short, the program first generates multivariate normal distributions for each correlated group of input variables, and then uses rank-preserving transformations from the normal distribution to the user-specified distribution for each input. Because the transformations preserve rank, and Spearman correlation depends only on rank, the Spearman correlations are unchanged by these transformations. Thus, the final distributions have the correct marginal forms and also satisfy the correlation targets.

When three or more variables are given pairwise correlation targets, it is possible that the combination of target correlations cannot be achieved. As a trivial example, if two variables have a correlation of one, then their correlations with any third variable must be identical. Even when none of the correlations are unity, impossible combinations may occur. There is a straightforward test (positive semidefiniteness) for the admissibility of the input correlation matrix that is automatically performed in SHEDS. There is also a slightly more stringent restriction (positive semidefiniteness of a derived matrix) that applies specifically to the NORTA method, and this is automatically tested by SHEDS as well. These tests are performed before any results are generated. If the tests are failed then an appropriate message is written to the log.

To specify correlations, in the bottom part of the interface screen “General Exposure and Dose Factors – Variability Distributions”, the user would check the box for “Correlate Input Variables” (see User Guide). A new screen lists all possible input variables that may be correlated. The user selects the subset of all variables subject to correlation. A second new screen appears, on which the user selects pairs from this subset along with the pairwise correlation target. Any pairs not assigned a target correlation are assumed to have a target correlation of zero. Note that this can lead to problematic situations. For example, if the user requests variables A, B, and C to be correlated, and gives large correlations to the pairs (A,B) and (A,C) while omitting the (B,C) pair, then the program may well fail, as it cannot generate a correlation of zero for (B,C), given the large correlations for the other cases.

When running SHEDS without the interface, a user can directly edit the Correlations data set to specify the desired correlations. Each record contains the names of the two variables in the pair, and a numeric variable for the size of the correlation. The user can enter as many pairs as desired. If the resulting correlation matrix is not positive semidefinite, a message is printed to the log during the model run. Note that "positive semidefinite" is a property reflected in the eigenvalues of the matrix (they are all non-negative); it does not mean that cells of the matrix cannot be negative. In general, negative correlations are acceptable, and both positive and negative correlations may appear in the same matrix.

.

3.2 General Issues

3.2.1 Units

Each quantity in SHEDS must be expressed in particular units. These are all metric, apart from standard time units such as hours and days. The exposure and dose variables all use micrograms (abbreviated “ug” in SAS code and file names) of the chemical agent as the basic unit. SHEDS operates as a mass-balance model in many ways, so it is useful to express quantities directly in mass units. Many exposure scientists use (mg/kg/day) for measuring exposures or doses. Such quantities are calculated as summary statistics from the SHEDS outputs, as discussed in a later section.

The units for the input variables are given in the Variables data set and are displayed by the interface. SHEDS does not explicitly recognize or manipulate these units, so it is not possible to change them. See Appendix F for a list of all inputs and their required units. The units on the “variables” data set are for informational purposes only, and the model would not recognize any changes even if the data sets were edited to have different units. A few input variables do not conform to the SHEDS standard that masses should be in micrograms. For example, the chemical application rate has units of (g/m^2), which is more familiar to most users. When necessary, the model internally converts such units into the preferred form of micrograms.

3.2.2 Formats

One advantage to using the user interface is that the user does not have to be concerned with data formats. When running SHEDS without the interface, all inputs apart from the command line arguments must be in the form of SAS data sets, and their entries must be formatted correctly. Those users who are very familiar with SAS should have little problem. It is strongly recommended that others employ the SHEDS interface.

Time series data in SHEDS are organized as columns in data sets. Each record is one point in time, and each column is one variable. The natural output of SHEDS is one record per diary event. When displaying or manipulating such time series, it is important to remember that the events are not of equal duration. In fact, the event duration itself is recorded as a time series variable in one of the columns on the data set. The calculation of time-averaged or time-integrated quantities such as daily totals will often require the explicit use of the event duration.

3.2.3 The default input files

The SHEDS installation package and the user interface come with default input files that can be changed by the user. While the data in these files has been deemed to be reasonable for the chemical and the scenarios featured, these data might not be appropriate for other model runs. It is the responsibility of the user to ensure that any default values that are retained are suitable for their purposes.

3.2.4 Simulation period

The simulation period in SHEDS is taken from the “real world” calendar, using the built-in SAS date functions. Therefore, a one-year SHEDS simulation may contain either 365 or 366 days, depending on which year is modeled. The days of the week also match the actual calendar, so January 1 is a Friday in 2010, but will be a Saturday in 2011, for example. The current version of SHEDS does not have any meteorological or ambient-air inputs, so the primary reason for selecting a particular year is to unambiguously assign a specific day of the week to each date. While a "generic" calendar could have been used, that would have negated the ability to use built-in SAS calendar functions.

All days in SHEDS are exactly 24 hours long, and all activity diaries from CHAD are also 24 hours (SHEDS does not simulate daylight savings time). However, the months have the correct number of days. Since months are not all of equal duration, SHEDS does not normally generate monthly exposure totals. However, for input purposes, SHEDS recognizes monthly chemical usage probabilities. Also, random variables may be resampled on a monthly basis.

The simulation period in SHEDS is an integer number of days. It can be as small as one, and can extend beyond one year. However, the latter is not recommended, as SHEDS does not alter a person’s age during the simulation. There is no restriction on simulations crossing from one calendar year to the next; for example, a 90-day winter season simulation could run from December 1 through February 28 of the next year.

Unless the time series option for specifying concentrations is used, the simulation period will be the same for all the individuals in a given SHEDS run. Under the time series option, the user supplies a set of concentration time series (one for each contact medium) at a set of houses. Since these inputs often come from field studies, the sampling dates may differ for the various houses. SHEDS randomly assigns each simulated individual to one of the houses. The simulation period then is set to match the dates for which concentrations are provided at that house. Therefore, all individuals assigned to the same house will share the same simulation period, but this may differ from the simulation period at another house, unless the input dates always match. This may become important when calculating summary statistics.

For purposes of determining chemical concentration via chemical usage, it is necessary to check for usage prior to the start of the simulation period. This is done to ensure realistic modeling of exposure throughout the simulation period, regardless of the start date. The chemical usage calendar in SHEDS may be extended back before “day 1” of the simulation. This is controlled by the "PriorUsage" setting, which gives the minimum number of possible usage days to consider. For example, the usage calendar might indicate that the house was sprayed on “day -22”. The concentrations resulting from these usages are determined normally, and the consequence is that the concentrations are often not zero at the start of the simulation period.

3.3 Job Settings

All of the variables listed under job settings in Table 3-1 are stored on the RunInfo file. The SHEDS interface automatically updates this file for the user, although advanced users may bypass the interface and edit the RunInfo file directly. The RunInfo file is stored in the top-level SHEDS installation directory, above the program code and default data sets. A single RunInfo file should suffice for all the SHEDS model runs performed on a given computer.

Each SHEDS model run requires a unique Run Name, up to 32 characters in length. Also, a Run Path is required. A folder or directory with the given Run Name is created at the location of the Run Path. Two model runs cannot share the same run name, even if located in different Run Paths. However,

an existing run may be deleted (to allow re-use of the Run Name), or the inputs to an existing run could be edited and the job re-run.

Each model run requires chemical names to be specified. The interface allows up to five chemicals in a single run, although the code allows an indefinite number. Each chemical name may be up to 32 characters and is used as SAS variable name. This means that letters, numbers, and underscores are permitted. Blanks and commas are not permitted and should be omitted or replaced by underscores. The chemical name cannot be missing or blank.

Each chemical has a metabolite name associated with it, which is also appended to certain SAS variable names. It is also limited to 32 characters, restricted to letters, numbers, and underscores. If the metabolite name is left blank, the code sets it to be the same as the chemical name.

The number of persons to simulate must be specified. This applies to a one-stage run, or to each inner (variability) step of a two-stage run. While a few hundred persons may be enough to locate the population mean within a few percent, more persons would be needed to reliably locate percentiles in the tail of the distribution. Several thousand persons are a better choice if computer resources permit.

The list of age-gender cohorts is required. The age groups are pre-defined and are displayed by the interface. For batch runs of SHEDS, these may be found on the "agegroups" data set. Different age groups may be selected for each gender. Every simulated person in SHEDS is assigned a specific year of age; the groups are partly a shortcut to avoid the need to specify each year separately. Also, age groups are used to pool the activity diaries into cohorts. This step has already been performed for all the groups on the default agegroups file. By altering this file and re-running the diary preprocessing, the age groupings used by SHEDS may be changed.

The start and end dates for the simulation must be expressed as SAS dates on the RunInfo file. This is automatically handled through the interface. If the user is editing the RunInfo file outside the interface, either use SAS programming, or else note that a SAS date field requires a numeric value, not a text string (for example, the entry appearing as '01JAN2010' must be entered as 18263, which is the SAS number for this date). The end date may equal the start date (a one-day run), but cannot be earlier than the start date.

The number of uncertainty iterations is only relevant for two-stage uncertainty-variability runs. The total number of simulations in a two-stage run is the product of the number of persons and the number of uncertainty iterations. For example, with 1000 persons and 100 uncertainty iterations, the program would create 100,000 simulated individuals, which would probably take several days.

The random number seed is important for reproducing an earlier run, or selected persons from an earlier run. The seed must be an integer between 1 and 2147483646 (SAS's upper limit), inclusive. There is no discernible relationship between runs that use different seeds. If no seed is present on the RunInfo file at the start of the job (or equivalently, the seed is zero), then the computer will generate a proper seed using the clock. This new seed will be recorded on the RunInfo file, to permit the run to be repeated later.

There are three ways to specify scenario usage in SHEDS: Fixed, Variable, or Time Series ('none' in the SAS code). The selected option is entered as a text string, but is not case-sensitive. The option 'none' requires that the user enter concentration time series directly (using the TimeSeries option for concentrations discussed below). Fixed usage means that all simulated persons share the same usage pattern, with the scenarios applied on the stipulated user-supplied dates. Variable usage means that each simulated person is randomly assigned usage dates according to the relevant rules and probabilities. Under variable usage, some persons might never use any of the given scenarios.

There are three primary options for the concentration method setting: Decay, Intervals, or TimeSeries (details are discussed in the Methods chapter of this manual). These are text strings but are not case-sensitive. The TimeSeries option requires UsageMethod=none; that is, no scenarios can be specified and no usage dates are assigned. Either Decay or Intervals may be used with either the Fixed or Variable usage options.

The dermal transfer method may be set to either TC (transfer coefficients) or to TE (transfer efficiency); this is a text string of length 2 and the variable is not case sensitive. The differences between TC and TE are detailed in the Methods chapter.

The method for assembling longitudinal activity diaries may be set to either 'eight' or 'new'. The eight diary method randomly selects two age/gender appropriate diaries for each season, one for weekdays and the other for weekend days. This was the only method available in SHEDS version 3. The 'new' method is based on the diversity (D) and autocorrelation (A) statistics, and requires the user to specify three additional run settings: the key diary variable, the D setting, and the A setting. Details of these two methods are discussed in the Methods chapter of this manual.

There are two settings that apply to Handlers. SHEDS only determines the exposure of one person at a time, so if someone other than the simulated person applies the chemical, the exposure during handling is not relevant and is not calculated. The simulated person can only be a Handler (that is, the applicator) if they are at least the minimum age specified for handlers. The other setting is simply called 'Handlers' and may be set to All, Some, or None. 'All' means that all persons over the minimum age are automatically handlers for all scenarios. 'Some' means that Handler status is randomly determined. 'None' means that no persons are handlers. With the 'Some' setting, a simulated person may become the handler for some scenarios but not others.

The maximum number of application dates for one scenario is a job setting that should normally have no importance; similar remarks apply to the maximum number of houses for use with TimeSeries input. These are defined here to bound loops and arrays in the code, but should be altered in the RunInfo file if there is any danger of exceeding the limits.

The variable controlling the fraction of diary events in multiple locations is primarily a modeling construct. Suppose one half of the house is treated, and the diary places the simulated person in the house for one hour. There are three possibilities: the full hour in spent in the treated half of the house, the full hour in the untreated half, or part of the time in each. This variable controls the relative frequency of each of these possibilities, subject to the constraint that over many diary events the average amount of time in the treated portion will equal the fraction of the house that is treated (one-half in this example). Thus, this variable does not alter the mean exposure, but can alter the pattern of exposure. The range of this variable is from zero to one. The default value of one-half may be the most consistent setting when comparing SHEDS version 4 to version 3.

SHEDS allows scenario usage before the start of the simulation period. Such prior usage leaves some residual chemical in the contact media at the start of the simulation period. Without prior usage, all media would start at zero chemical concentration and remain there until the first scenario usage event, which may be well into the simulation period. This is especially important with the Variable-Dates usage method. It cannot be used with TimeSeries input at all, and with Fixed-Dates one can always arrange to have scenario usage occur at the start of the simulation period. Setting the prior usage period to 30 days should be sufficient, since usages more than 30 days before the simulation begins will likely leave very little chemical to be contacted. Note that when the prior usage period extends into the previous calendar year, the usage patterns from the end of the current year are repeated.

NOAEL stands for 'no observed adverse effect level' and is used to set MOEs (margins of exposure). There are three NOAEL values, for the dermal, inhalation, and ingestion pathways. For the

latter two, the MOE is obtained by dividing the NOAEL by the daily dose, which is the amount of chemical crossing the exposure surface per day. All NOAEL values in SHEDS are assumed to be in units of (mg/kg/day). The dermal NOAEL refers to the amount of chemical on the skin surface (the exposure), whereas the ingestion and and inhalation NOAELs refer to the amount of chemical entering the body (the doses).

SHEDS can retain the full set of event-level exposures and related variables for selected persons, as well as the usual daily summaries. This is not practical for all persons in a large run. This setting is called the *person_numbers* input and is normally used when repeating an earlier run and the user wants more information on selected persons. The person numbers are written in a space-separated list. SHEDS will produce output only for the persons on the list, unless the list is empty.

The remaining job settings are flags or switches that control job behavior. Setting allout=1 adds a lot of intermediate variables to the output data sets, including the settings of all the random inputs. This is necessary when correlating or regressing output data on inputs for sensitivity analysis, but may be useful in other cases as well. Allout=0 produces smaller data sets.

Setting debug=0 produces almost no messages on the SAS log, which is necessary when running a job with thousands of persons (otherwise, the log would fill up and halt job execution). Setting debug=1 is useful for following program flow. Debug=2 is also available and writes even more information to the log. These options should only be used with very few persons (say, 10 or fewer).

Setting DiaryPrep=1 makes SHEDS pre-process the activity diary database prior to starting the model run. This step converts the ASCII files extracted from CHAD into the SAS data sets used by SHEDS. Since these SAS data sets are permanently saved and are part of the default inputs, the diary pre-processing would only be necessary if the original diary files are modified, perhaps by adding or removing diaries.

Setting SensPct=1 allows a SHEDS sensitivity analysis using the percentile scaling approach. The variables to be analyzed should also be marked on the Variables input file by setting SensPct=1. Variables with SensPct=0 will be left at their median values. If J variables are marked for analysis, then SHEDS will perform (2J+1) passes through the inner (variability) loop. This can be used to compute job execution time. For example, with J=7 and N=1000, the program will simulate 15,000 persons and take the same time as a variability run of that size.

Setting SensSobol=1 allows conducting a Sobol sensitivity analysis. This requires that each record on the Variables input file be assigned a group number, under the SobolGroup column. Groups should be sequentially numbered (from 1 to J, the total number of groups). The collective sensitivity of the model to each group will be assessed. With J groups, a total of (2J+2) passes through the inner loop are required, which can be used to estimate the execution time.

Setting SaveLog=1 saves the log screen to the permanent data set 'saslog'txt' in the output folder for the run. With SaveLog=0 the log appears on the screen only and is erased when the SAS session ends.

Setting Progress=1 causes SHEDS to write one line on the log each time a person is completed. For small to medium runs this is a useful indicator of when the job may be expected to finish. For extremely large runs, even this indicator may cause the log window to fill prematurely, so it can be turned off with Progress=0.

Setting CorrInputs=1 causes SHEDS to read the input file containing target correlations for input variables. With CorrInputs=0 this file is not used and no correlations are made.

Setting ForPBPK=1 causes the event-level output file intended for input to a PBPK model to be produced. This file can be very large (about 1 megabyte for each simulated person), even though it only

contains a small number of variables. Setting forPBPK=0 suppresses this file, which saves a lot of file space.

3.4 Scenario-specific inputs

SHEDS requires dates, times, and amounts of usage, or the rules for selecting these variables, unless the TimeSeries input option is used. Scenario usage directly affects the determination of the chemical concentration time series. There is no necessity in SHEDS to have usage occur at the start of the simulation period.

There are 9 scenarios currently defined on the SHEDS default input files (indoor crack and crevice aerosol, indoor crack and crevice liquid, indoor fogger, indoor flying insect killer, outdoor lawn granular, outdoor lawn liquid, outdoor vegetable garden powder, pet spot-on, pet liquid). The default scenarios are intended to represent the major pesticide use scenarios; other default scenarios (e.g., repellents, impregnated clothing, etc.) could be added in the future. The user selects which of these scenarios are to be used in a given model run, from any single one up to all ten. Others may be added to this list without needing to alter the SHEDS code, although the GUI does not currently support this option. To add more scenarios, the user must edit the relevant input files in a consistent manner. The "Scenarios" input file has one record for each scenario. Each scenario is given a unique name (up to 12 characters, without spaces) and a location (one of Lawn, Garden, Indoor, or Pet). Setting Use=1 on the record for a scenario allows it to be considered in the model run (setting Use=0 effectively eliminates that scenario from the model).

There are six other files that mention specific scenarios and may need to be updated when new ones are defined: the "Handlers", "Decay", "Intervals", "Dates-fixed", "Dates-variable", and "Variables" files. Examination of the existing versions of these files should provide guidance on the information needed. The Dates-fixed and Dates-variable files have one record for each scenario. The Handlers file has three records for each scenario. The Decay and Intervals files have variable numbers of records for each scenario, depending on the location and the media affected. The default files provide examples for each location, so new scenarios should follow the same rules. The Variables file lists the full set of variables for which random numbers must be drawn: there are three such variables specific to each scenario.

The list of relevant chemical agents found in each scenario is required. In a multi-chemical run, not all scenarios will necessarily contain all chemicals. When scenario usage occurs, all of the chemicals found in that scenario will have their concentrations elevated simultaneously. See the section below on scenario-chemical inputs for more details.

The probability that the simulated person is the handler or applicator is needed for each scenario. This probability only applies to persons who meet the minimum age requirement. SHEDS does not specifically model professional applicators. If there are two adults in a given home, the chance that the one being simulated is the handler should be no more than one-half (it may be less because an outsider may be the handler). The handler probability is scenario-specific, as some scenarios are commonly applied by the homeowner, whereas others are usually professionally applied.

A new scenario-specific variable in version 4 is the fraction of the house area that is treated, which is for indoor scenarios only. Some are typically used only in a small area, whereas others can cover the entire house. This directly affects the probability of being the in treated area while indoors. SHEDS assumes maximum overlap in treatment areas. For example, if one scenario covers 20% of the house and another covers 50%, then 20% of the house is covered by both, 30% by only the second, and 50% by neither.

3.4.1 Inputs for the Fixed-Dates option

There are two methods of specifying application scenario dates (Fixed-Dates and Variable-Dates), and each has a list of variables needed for each scenario. Note that the choice of Fixed-Dates or Variable-Dates is made once per run and applies to all scenarios (and all persons) in that run.

Under the Fixed-Dates option, for each scenario in the model run, the user must enter the list of usage dates. The dates are indicated by sequential day numbers, relative to the start of the simulation. Thus, day 1 is the start date of the simulation period, day 2 is the day after the start date, and so on. The user may include zero or negative numbers in the list. Thus, the list (-6,0,15,90) indicates that usage occurs seven days before the start of the simulation, again on day zero , a third time on day 15, and finally on day 90. If the simulation were (say) only 60 days long, then the final usage in this example would be irrelevant (having no effect on exposure or dose during the simulation). The hour of usage is also entered for each scenario.

It is common in Fixed-Dates runs to apply the chemical at the start of the simulation. This can be achieved by setting the application date list to (1), but then the first day could include some time before the application actually occurs. This might have implications for the exposure distribution on day one. If the user prefers to have the potential for a full day of exposure on day one, it might be better to schedule the application on day zero, which is the day right before the start of the simulation period.

For each scenario used in a Fixed-Dates run, the reentry time may be specified. This is a fixed number of hours and will be the same for all persons in the run. The variables p_lawn, p_garden, and p_pet are not relevant. Every simulated person shares the same set of application dates. Therefore, if a lawn scenario is selected for the run, all persons in the run are assumed to have a lawn; the same approach applies for gardens and pets.

3.4.2 Inputs for the Variable-Dates option

The Variable-Dates option considers the population variability in scenario usage. The SHEDS user specifies the list of scenarios to be considered in this run. Each simulated individual will randomly be assigned usage patterns, based on the user-supplied probabilities. It is likely that very few individuals will use all scenarios, and in fact, some individuals may use none of the specified scenarios.

Scenarios associated with the Lawn, Garden, or Pet locations may only be used by persons who have those. For example, if p_Lawn = 0.75 then 25% of the simulated people will not have lawns at their homes and will not be allowed to use lawn scenarios. Indoor scenarios may be used by anyone. Note that "use" in SHEDS refers to the home, not the individual. Obviously, a baby would not personally apply chemicals, but would still be considered to be a "user" if the scenario is used in that home. The term "handler" refers to a person who actually applies the scenario and may become exposed during the application process. Children below the user-specified cutoff age are never handlers.

For each scenario the overall probability of use must be provided, which assumes that the person meets the requirements. For example, suppose the probability of having a lawn is 80%, and the probability of treating the lawn with the specified scenario is set to 50%. Then in a model run of 1000 persons, about 400 should have one or more lawn treatments (that is, about half of the 800 or so who have a lawn). The number is not fixed because usage is randomly determined for each person.

In addition, three probability vectors are required for each selected application type:

1. Pr (#usages per year)
2. Pr (month of usage)
3. Pr (day of week)

The interface allows up to 12 usages per year (to keep data entry into the GUI convenient for the user); more could be specified if the input files are edited outside the interface. Each probability in the list corresponds to a number of usages equal to the position in the list, so the first probability is the likelihood of exactly one usage per year. Since the sum of these probabilities should be one, the implication is that each probability is conditional on at least one usage occurring (that is, non-users are excluded). The probability of zero usages is not included on this vector because it may depend on other factors; for example, if the simulated person does not have a garden then the probability of zero usages of a garden scenario is one. SHEDS first determines whether a person is a user or not, and in the former case randomly selects the number of usages using this probability vector, with one usage being the minimum allowed result. For a simulation period less than a year, this probability vector still represents a full year. The model may select application dates at any point in the calendar year, which could be before, during, or after the simulation period.

The probability vector for the month also has 12 slots; the first represents the fraction of all applications of that type that occur in January, the second corresponds to February, and so on. Each scenario has its own vector, as some scenarios are used year-round while others are seasonal. The probability vector for the day of week has 7 slots, with the first representing the fraction of all applications of that type that occur on a Sunday, the second on Monday, up to the seventh which represents Saturday.

Three other probability vectors are needed for the Variable-Dates option, for selecting the hour of the day at which the usage occurs. These vectors are specific to location rather than to scenarios. There is one TimeOfUse vector for indoor scenarios, one for outdoor (that is, Lawn or Garden) scenarios, and one for pet scenarios. These vectors have 12 slots, with the first representing 6 a.m., the next 7 a.m., up to the twelfth slot representing 5 p.m..

3.4.2.1 Re-entry prohibition times

The re-entry prohibition or ban periods are also discussed in the Methods section. They represent short periods immediately after usage when people are not permitted to enter a treated area. The Variable-Dates method uses three: one for indoor scenarios, one for outdoor (lawn and garden) scenarios, and one for pet scenarios. The re-entry prohibition is a whole number of hours from the time of usage. Any type of distribution may be used. The code automatically truncates any non-integer part of the result to produce a whole number of hours.

The prohibition period represents the minimum possible time until post-application exposure occurs. The activity diary must indicate the correct location for contact to occur. For example, when a lawn scenario application occurs, contact must wait until the diary indicates an out-home event. It can happen that no appropriate event occurs on the diaries for a long time, even if the prohibition period is short.

3.4.3 Co-Occurrence Factors

When the Variable-Dates option is used, the default assumption is that the dates of chemical usage are selected independently. To change this assumption, the user should specify co-occurrence factors.

For a model run with a single scenario, only a blackout period can be imposed. This prevents two or more applications of the same scenario from occurring closer together in time than the minimum allowed by the blackout duration. For example, if an application occurs on day 100 of the simulation and the blackout period is set to 7, then another application of the same scenario could occur on or before day 93, or on or after day 107, but not between those dates. With multiple scenarios, a separate blackout period applies to each one. The minimum blackout period is one day, which means that the same scenario cannot be used twice by the same person on the same day. As with all variables, blackout periods apply only to the person currently being modeled; there are no cross-person influences in SHEDS since every simulated individual is independent of all others.

When two or more scenarios are used with the Variable-Dates option, then usage of one may affect the usage probabilities for the other(s). Each scenario is assigned a priority which indicates the order of evaluation. For a given person, all usages dates for one scenario are set before any usages dates are selected for the next. For each scenario the usage probabilities are based on the probability vectors for number of usages per year, month of use, and weekday of use. These may be modified on days close to those on which another scenario is used, based on three user-supplied inputs. One is the list of the other scenarios that affect the probabilities for the current scenario. A second input is the width of the adjustment window (the number of days on each side of the usage day which will have their probabilities modified). The third input is the multiplicative factor for adjusting the probabilities. If this factor is over one, then the current scenario becomes more likely to be used on or near the usage dates for the others. Conversely, if the factor is less than one then the current scenario becomes less likely to be used when the others are used. A multiplicative factor of zero creates an outright ban on usage during the window period. Figure 2-2 and Figure 2-3 (shown earlier) illustrate these adjustments.

3.5 Inputs that are scenario-specific and chemical-specific

Each of the variables discussed in this section requires one distribution for each scenario-chemical combination that is included in the model run. If certain combinations do not exist (because the scenarios do not necessarily each contain all the chemicals being modeled), then the variables for that combination are set to zero (this is handled automatically by the interface).

The chemical application rate is the amount of active ingredient (chemical agent) applied, divided by the treatment area. The units are (g/m^2). This information may be obtainable from product labels and is used to estimate exposure during handling (the application process). For example, suppose a container holds 100 grams of a mix containing 2% active ingredient, and is applied to an area of 50 m^2. The application rate is then 100 * .02 /50 = 0.04 g/m^2. Note that this is *not* the same as the "initial surface concentration" which only considers the amount available for potential transfer.

The unit dermal exposure for handlers is essentially the fraction of the applied chemical mass that ends up on the applicator's skin. The units are (µg/g) or parts per million by mass. Suppose in the above example the unit dermal exposure is 10 ug/g. Then the amount getting on the skin of the handler would be 100 grams product applied * 0.02 (mass fraction ai in product) * 10 µg ai exposure/gram ai applied = 20 µg ai exposure. This base amount would be adjusted for pre-existing loading and for amount of clothing. Since SHEDS keeps separate totals for the hands, the dermal exposure is divided into two terms, with a fraction F_DermalHands going to the hands and the remainder going to the rest of the body.

The unit inhalation factor for handlers is also measured in (μg/g) and represents the fraction of chemical mass applied that ends up being inhaled. Suppose this factor is 0.3 μg/g. Then the amount inhaled in the above example would be 100 * 0.02 * 0.3 = 0.6 μg. In SHEDS version 4 the amount of chemical inhaled is called the inhalation dose. The inhalation exposure is actually the air concentration, which is not determined in this example.

For indoor scenarios, two initial surface concentrations are needed: on hard floors and soft floors (usually called "carpets" in SHEDS). These are measured in (ug/cm^2) and represent the amount of chemical loading that may easily be dislodged from the surface. In a field study, the amount of chemical picked up by a wipe sample or perhaps a PUF roller would be suitable as the initial surface concentration. Lawn, garden, and pet scenarios each require one initial surface concentration, but the meaning is essentially the same, as it represents the amount of chemical loading that is easily removable or available for transfer. Indoor scenarios also have an initial concentration in air with units of [ug/m^3]. Outdoor scenarios have a unit inhalation dose during handling, but it is assumed that any air concentration would have dissipated by the time any post-application exposure occurs.

The user may also specify an initial concentration in dust (for indoor scenarios) or in soil (for outdoor scenarios). The units for either are (ug/g). SHEDS uses the indoor dust loadings to convert the units to (ug/cm2), whereupon it is combined with other forms of surface loading to produce the total chemical mass per unit surface area. The user may specify either the chemical surface loading or the concentration in dust, but note that if both are used then SHEDS will sum the resulting loadings.

Outdoors, the concentration in soil may be specified, similar to the concentration in dust. Since outdoor soil loadings (thicknesses) are not well defined, SHEDS computes this starting with the daily soil ingestion rate for a 4-year-old child. SHEDS then uses the other input variables for this age to calculate the soil loading that would produce such a soil ingestion rate. Details of this calculation are given in the comments to the %SoilProperties macro in the SAS code. This soil loading is applicable regardless of the age of the simulated individual.

For the Intervals method, all of the initial surface concentration variables are repeated four times; for the day of application, 1-7 days later, 8-30 days after application, and 31 or more days after. For the Decay method, only the initial (day of application) distributions are needed.

All of the above variables (except for soil ingestion and soil loading) are automatically correlated for a single scenario (but not across scenarios). This is achieved by selecting the same percentile from each distribution. Thus, if the chemical application rate happens to be at the 70^{th} percentile for a given application, then so will be the initial surface concentration in the carpet, the hard floor, the dust, and in the air. This is reasonable since all these variables share a causal relationship (the user applied the chemical either sparingly or heavily).

For the decay/dispersion method the user must also specify the fraction of chemical lost per day. This is not the same value as the decay constant in an exponential decay equation, nor is it the half-life. For many pesticides between 5% and 20% will be lost per day (these would be entered as 0.05 or 0.20, respectively).

Both the decay/dispersion method and the intervals method require the variable f_UTratio, which measures the ratio of untreated/treated concentrations. This ratio can be estimated from a fugacity model or measurements studies, and should always be less than one. Four distributions for f_UTratio are specified for the intervals method (for day 0, days 1-7, days 8-30, and days 31+).

For the decay/dispersion method the variable f_UTratio has a slightly different meaning. It represents the ratio at the point in time when the untreated area reaches its maximum concentration. The treated area reaches its maximum concentration immediately after an application and thereafter declines monotonically. Assuming no prior usage, the untreated area starts at zero concentration and slowly builds

up as chemical disperses out of the treated area. But as the treated area loses chemical, the flow to the untreated area slows down and eventually cannot make up for the chemical losses in the untreated area. From then on, the chemical concentrations decline in both areas. However, the concentration ratio continues to increase, slowly moving toward equality. If the f_UTratio =0.05, that means that when the untreated area reached peak concentration, it was at 5% of the concentration in the treated area (as measured at the same time; this is not 5% of the initial concentration). SHEDS handles the rest of the calculations needed to achieve this result. All media except air share the same f_UTratio value.

3.6 Activity Diary-Related Inputs

3.6.1 CHAD Diaries

SHEDS selects its activity diaries from EPA's Consolidated Human Activity Database (CHAD), available at http://www.epa.gov/chadnet1 (see McCurdy et al., 2000). CHAD originally contained over 22,000 one-day (24 hour) activity diaries, consolidated from more than a dozen original studies. Recently, some new studies have been added and there are now around 34,000 diaries in CHAD. Other sources of activity diaries could potentially be used in SHEDS, but this would require those diaries to first be reformatted to follow the CHAD coding conventions.

Each CHAD diary represents a 24-hour (midnight-to-midnight) time sequence of activities and locations for a particular individual. Each such diary is assigned a unique CHAD ID. The age, gender, calendar date and certain other properties of the individual are noted. The diary consists of a series of records, each one representing one "diary event." Each diary event has a start time, stop time, duration, activity code, and location code. All times are recorded to the nearest minute, and the stop time for one event must equal the start time for the next event (no time gaps or time overlaps are permitted). If the activity diary originally had a time gap, then in CHAD a dummy activity is generated to fill the gap, with activity and location set to missing. SHEDS handles such cases by setting the activity and location to equal those of the prior event. Each diary event lasts until the activity or location changes, or until the next clock hour starts. Since no event may cross into the next clock hour, the maximum duration of a diary event is 60 minutes. A CHAD diary has a minimum of 24 diary events, and may have up to 100 or more. The mean number of events in CHAD is about 40 per diary.

A few of the studies that were consolidated into CHAD recorded more than one diary day from a given individual. In such cases, the overall diary is broken into separate 24-hour diary days. Multiple diary days from the same individual can be identified since they generally share the first 8 characters of the "CHADID" variable, with only the final letter varying. However, SHEDS does not attempt to keep these pieces together; one may be selected without the others.

The default directory that comes with the SHEDS interface contains copies of the CHAD activity diaries, already converted into SAS format. These are the files "diaryquest.sas7bdat," "diaryevents.sas7bdat," and "diaryevents.sas7bndx." The first two of these contain the data while the third indexes the diaries. If these files are not present or a later version of CHAD is obtained, these must be reconstructed. Also, if the diary pre-processing options are to be changed then the raw diary files (.txt extension) must be obtained. The diary pre-processing can only be run outside of the user interface.

3.6.2 Pre-processing the CHAD files

The SHEDS installation package already contains the CHAD data needed for SHEDS. If CHAD is updated then the user may wish to re-extract the data for use in SHEDS. Rules to do this were provided with SHEDS version 3. However, the main CHAD database (the file "chad2000_v1047.mdb") only contains the original 22,000 CHAD diaries and has not been updated with the additional 12,000

diaries recently put in CHAD format. Instead, SAS files containing the relevant information needed for SHEDS are available for the larger database; these are called "chad_questionnaire.sas7bdat" and "chad_events.sas7bdat."

The SHEDS diary preprocessing module %DiaryPreProcess will filter these diaries using the diary selection/rejection rules and format the allowed diaries into cohorts and diary pools suitable for use by SHEDS. This macro must be run if the diary database changes (for example, if new diaries are added), or if the user wants to alter the acceptance/rejection rules. Otherwise, there is no need to run the diary pre-processing module, as this has already been done in the SHEDS installation package.

At present, the diary pre-processing cannot be done via the user interface. Apart from the events and questionnaire files, the pre-processing requires the file "DiaryQARules.sas7bdat," which contains instructions for rejecting certain diaries for use in SHEDS. The "DiaryQARules" file supplied with the installation package contains the rules given in Table 3-5.

Table 3-5. Default rules for diary pre-processing in SHEDS

Test	Interpretation
(gender) in ('M','F')	Gender of diary respondent must be known
(age>=0) and (age<=99)	Age must be known and not over 99
LocUX<120	Less than 120 minutes in unknown locations.
ActUX<360	Less than 360 minutes in unknown activities.
Pai<4	Daily average METS must be less than 4 times basal rate
Qcactloc<=240	Must not be more than 240 minutes of unusual location-activity combinations
Qceattime<=360	Must not be more than 360 minutes spent on meals

These rules may be modified as needed. The restriction on ActUX was only 120 minutes in SHEDS version 3, but a large number of new diaries were added from a children's study on ozone averting behavior, and many of these diaries have missing data after 8 p.m. Since children are often in bed soon after this time, it may be preferable to set the rules to permit these diaries to be included rather than to lose the information on them regarding activities earlier in the day. In any case, location plays a greater role than activity in determining the SHEDS micro, and the limit on missing location time is just 120 minutes. The last three rules eliminate some known problem diaries from being used.

A diary that fails any of these tests is removed from consideration for use in SHEDS. All other CHAD diaries are acceptable. Appendix D contains summary information and diary counts for CHAD diaries used in SHEDS.

The list of criteria may be modified or expanded. If a rule contains any logical connectives like 'OR' or 'NOT', then the terms on either side of the logical operator should be put in parentheses, for example: (gender='F')or(gender='M'). The reason for this is that blank spaces are not permitted inside a rule, and without the parentheses the operator would not be parsed correctly. The diary QA rules may be used to limit the pool of available diaries. For example, if the model is being run only on children ages 1-6 years, then one could add the QA rule (1<=age<=6), to limit the diaries to the correct age range.

Similarly, one can require that all diaries have some outdoor time by using the rule (outside>0). Restricting the pool of diaries results in both space and execution time efficiencies. Any variable on the

dataset RawQuest2 can be used in a QA rule. These include both variables from the CHAD questionnaire file and some additional variables like sums of time in various locations that are added in the 'diarysums' step. Note that every micro has its own variable of the same name, which contains the number of minutes spent in that micro. SHEDS uses the following five micros: in_home, in_veh, in_oth, out_home, and out_oth, in addition to the CHAD variables. Table 3-6 shows a complete list of variables that may be used in QA rules:

Table 3-6. Variables that may be used in diary QA rules.

Variable	Meaning
actUX	total time (minutes) in CHADact 'U' or 'X'
age	age in full years
age1	age as in CHAD (decimal years for ages below 2)
airCond	Y=has air conditioning at home, N=no, X=unknown
asthma	Y=has asthma, N=does not have asthma, X=unknown
avgtemp	daily average temperature (Fahrenheit)
bath	total time (minutes) spent in bath/shower
chadid	unique ID for each CHAD diary
county	county of residence
date	SAS date of diary day
day1	day of month (range 1-31, unless missing)
dayNum	day sequence number for this individual
educat	2-character code for educational level
employed	Y=employed, N=not employed, X=unknown
fuel	1-character code for type of heating fuel
fulltime	Y=employed full time, N=not fulltime, X=unknown
garage	Y=has a garage, N=no garage, X=unknown
gasstove	Y=has a gas stove, N=no gas stove, X=unknown
gender	M=male, F=female, X=unknown
heartlung	Y=has a heart or lung condition, N=no, X=unknown
heating	2-character code for system for home heating
hourrain	hours of rainfall (0-24) on diary day
housing	2-character code for type of housing
in_home	total time (minutes) in in_home micro
in_oth	total time (minutes) in in_oth micro
in_veh	total time (minutes) in in_veh micro

Variable	Meaning
inchrain	inches of rainfall on diary day (to 0.01 inch)
income	household annual income category (see CHAD)
jobhours	hours worked at job in prior week (see CHAD)
locUX	total time (minutes) in CHADloc 'U' or 'X'
maxtemp	daily maximum temperature (Fahrenheit)
month1	month of year (range 1-12, unless missing)
nDays	number of diary days in CHAD from this individual
nomets	total time with missing 'metslink' variable
nomicro	total time with missing (blank) micro assigned
occup	occupational category, X=(unknown or not working)
out_home	total time (minutes) in out_home micro
out_oth	total time (minutes) in out_oth micro
outside	total time (minutes) in any outdoor micro
pesticide	Y=used pesticides recently, N=no, X=unknown
pid	personal ID from original study
qaflag	# of failed QA test (qaflag=0 means it passed)
qcActLoc	CHAD flag for activity-location mismatch
qcEatime	CHAD flag for minutes spent eating
qcHeavy	CHAD flag for minutes with heavy breathing
qcInfer	CHAD flag for minutes in inferred diary events
qcLong	CHAD flag for most minutes in same act and loc
qcMeals	CHAD flag for number of meals
qcMetab	CHAD flag for minutes at high METS (>3.5)
qcMiss	CHAD flag for minutes with missing act and loc
qcSleep	CHAD flag for hours spent sleeping
qftravel	1=mismatch between morning and evening travel, 0=OK
race	A=Asian,B=black,H=Hispanic,O=other,W=white,X=unknown
reccount	# of event records on diary day
school	total time (minutes) spent at school
season	P=spring, S=summer, F=fall, W=winter, X=unknown
sleep	total time (minutes) spent sleeping
smoker	Y=person is smoker, N=not smoker, X=unknown
smoker2	Y=person lives with smoker, N=no, X=unknown

Variable	Meaning
state	20-character political state of residence
student	Y=is a student, N=not a student, X=unknown
travel	total time (minutes) spent in travel
wdwe	WD=Monday-Friday, WE=Saturday or Sunday
weekday1	3-character day of week (SUN,MON,TUE,etc.)
weekend	1=Saturday or Sunday, 0=other day of week
weight	person's weight in kilograms
work	total time (minutes) spent working at job
wraptime	actual start time for diary day
year1	4-character year for diary day
zipcode	5-character zip code, X=unknown

Some of the above variables such as ‘fuel,’ ‘heating,’ and ‘occup’ have codes or settings that are not obvious; these can be found in the CHAD documentation if such variables are of interest.

The use of these rules can provide the user with control over the pools of acceptable diaries. However, at present this feature is not available through the user interface. If the rules are changed, then the %DiaryPreProcessing macro must be invoked, either before or during the SHEDS model run.

3.7 Concentration-Related Inputs

As discussed above, SHEDS has three options for specifying concentrations: decay-dispersion, intervals, and chemical time-series input. Each of these is described in a subsection below. The first two of these methods are scenario-specific, so chemical concentrations can only exist if the given scenario was applied at the residence of the simulated individual. SHEDS also permits an outdoor background concentration on lawns and gardens, not connected to any scenario, as described below.

The object-floor concentration ratio is used to calculate the chemical concentration on the objects being mouthed. This is expressed as a ratio or fraction of the concentration on floor surfaces in the same location. The assumption is that objects such as toys were not directly treated, but were placed on treated surfaces later on. Some of the chemical on the floor will then be transferred onto these objects. This variable is unitless and is sampled once per person.

3.7.1 Background Concentrations

A "background" concentration refers to one that is not accounted for by the application schedule for the specified scenarios. SHEDS restricts the definition of background to chemical that results from usage at other houses (or non-residential usage). Chemical that is present at the start of the simulation period that has resulted from prior chemical usage is modeled in SHEDS, based on the usage patterns assigned to that house.

The outdoor background in SHEDS may be assumed to have drifted from a neighbor's yard. This background allows people without any chemical usage at their homes to nevertheless acquire some exposure. Otherwise, all persons who are non-users of the modeled scenarios would have exposures of zero at all times. The background is not used with the timeseries approach since that is meant to

represent total chemical concentration from all sources, including background. The user may set the background to a point value of zero if it is not desired in a particular run. In SHEDS version 4, the background is sampled once per person and does not change during the simulation; it is intended to reflect the proclivity of the neighborhood to use the chemical agent. While the background should be time-varying, it would be expected to have a strong seasonal and weather-based pattern. Modeling this variation is beyond the scope of SHEDS at the present time.

No indoor background is included in SHEDS version 4, apart from the chemical left over from previous indoor applications. If indoor background were modeled, it would result either from track-in or via deposition from infiltrated air. Since a major source of this would be the person's own yard, if it were treated, there should be a relationship between the outdoor concentration and the indoor background for each simulated person. However, algorithms to model this relationship have not been developed for inclusion in SHEDS.

SHEDS might have requested that the user input initial concentrations, but it was decided that it was better to model it based on each person's known chemical usage patterns. SHEDS already calculates chemical usage for all months of the year, even if the simulation period is shorter, since the usage inputs are on an annual basis. In version 4 the user specifies one additional input which is the number of days prior to the simulation period during which chemical usage should be considered. Suppose this is set to 30 days. Then any usage in the 30 days right before the simulation begins is also modeled, to determine the amount of chemical present at the start of the simulation. No exposure is calculated during this phase, since no activity diary has been generated. If the window (which is 30 days in this example) extends back into the previous calendar year, then the usage calendar for the end of the current year is used (that is, the pattern for December of this year is copied onto last year). The result of this calculation is that the initial chemical concentrations in each house are consistent with the typical chemical usage patterns established for that house. Furthermore, if the primary source of indoor "background" is due to prior indoor usage, then version 4 includes this background. If the user does not want to include this effect, it may be disabled by setting the window to zero days, in which case all chemical concentrations start the simulation at zero.

3.7.2 Decay-dispersion method

3.7.2.1 Indoor scenarios

When using the decay-dispersion method, each indoor scenario has seven inputs per chemical. The first three are the initial concentrations on hard floors, on carpets, and in dust. The dust concentration may be set to zero if it is implicitly included in the other two. The next two inputs are the air concentration in the first hour after application and at 24 hours. The final two inputs are the decay rate and the untreated/treated concentration ratio. The concentration units are ($\mu g/cm^2$) for the hard and soft surfaces, the units are ($\mu g/g$) for dust, and are ($\mu g/m^3$) for air. The decay rate is common to all four media and is expressed as the fraction lost per day. Thus, a rate of 0.05 implies that the chemical concentrations are reduced by 5% with each passing day. The final variable is the untreated/treated concentration ratio, which requires some explanation.

The difficulty arises from the fact that the concentration ratio between the untreated and treated rooms is not constant, but starts very small and generally increases (toward equality) with time. Suppose both the treated and untreated rooms are initially at zero concentration and an application occurs. The concentration in the treated room jumps to some initial value and then follows a first-order exponential decay over time, using the distributions discussed above. After this application, the chemical starts to disperse into the untreated room. The concentration in the untreated room continues to increase for several days, but it is also subject to decay (using the same rate constant as for the treated room). As the

concentration in the treated room drops, so does the amount of chemical entering the untreated room. Eventually, the rate at which chemical enters the untreated room drops below the rate at which it decays, and the concentration in the untreated room starts to drop. From this point on, the concentrations in both rooms continue to decline towards zero. Thus, there is a unique point in time when the untreated room is at maximum concentration. The untreated/treated concentration ratio in SHEDS is defined to be the ratio at the time that the untreated concentration reaches its peak.

The exact relationship between the treated and untreated rooms is quite complicated, depending on room geometry, size, flow rates, and other factors. A proper analysis would require a detailed indoor mass-balance model for the chemical. The simpler decay-dispersion logic in SHEDS assumes that the variables affecting flow rates are fixed and that the flow is proportional to the concentration difference. In a practical monitoring situation, it should be obvious when the untreated room reaches maximum concentration (assuming fairly regular monitoring), and the untreated/treated ratio simply requires taking two concentrations measurements at this time. This should be easier to measure that actual chemical flow rates between the rooms. It may be shown that the flow rate may be uniquely determined, given the decay rate and the untreated/treated ratio used by SHEDS. This logic has been programmed into the SHEDS code.

The initial concentration in dust is a required input for indoor scenarios, but it may be set to zero. After application, SHEDS does not distinguish chemical in dust from other forms. The initial chemical concentration in dust is expressed in units of ug/g, or micrograms of chemical per gram of dust. This may be obtained by lab analysis of vacuumed dust samples, for example. SHEDS also has an input for the amount of dust on the floor, measured in ug/cm2. This latter input may have different distributions for hard floors and carpet; each is sampled once per house and is common to all indoor scenarios, unlike the chemical loading in the dust. SHEDS translates the loading in dust to an effective loading per unit floor area and then adds this to any other initial floor loading (not in dust). When specifying the initial concentration on hard floors (in $\mu g/cm^2$), on carpet (in $\mu g/cm^2$) and in dust (in ug/g), remember that the last of these is added to the first two, so the user should ensure that double counting of the same chemical mass does not occur.

3.7.2.2 Outdoor and pet scenarios

The lawn and vegetable garden scenarios each affect one surface medium. For each scenario three distributions are needed, one for initial surface loadings in μg/cm2, one for initial chemical concentration in soil in ug/g, and one for the fraction lost per day. To avoid double counting, the concentration in soil should be set to zero if the initial surface loading represents all of the applied chemical, or if it refers to the dislodgeable portion but includes all phases of the chemical. On the other hand, if there are separate concentration data for various phases of the chemical such as residue and soil (which may be the case in a measurement study), then a non-zero soil concentration should be supplied to SHEDS. While the chemical may disperse outdoors, it is assumed that the concentrations in non-treated areas are negligible. Therefore, the UTratio variables are not needed.

As discussed above, SHEDS also asks the user for the daily soil ingestion rate. The code has a built-in macro for determining the effective soil thickness that would produce this rate, found empirically by randomly sampling the variables in the transfer equations. If the user supplies initial chemical concentrations in soil, these are converted to effective surface loadings in μg/cm2, using the calculated soil layer thickness. These are then added to any initial concentrations specified in μg/cm2, as is the case for indoor scenarios.

Each pet scenario affects only one contact medium. For purposes of contact, the pet is effectively another contact medium. Unlike the other contact media, the pet may be contacted in the treated room, the untreated room, or outdoors. The pet carries its own concentration with it, and does not adjust to its

surroundings. For each scenario affecting pets, a distribution for initial concentration in ug/cm2 is required, and another distribution for the decay rate is required.

3.7.3 Intervals method

For each combination of scenario and affected medium there are four inputs required for the intervals method. All four are distributions for concentrations, differing only in the time at which they take effect. The first applies on the day of the application. The second distribution is used from one to seven days after the application date. The third covers the period from eight to thirty days after application, and the fourth is used for all longer intervals.

Indoor (non-pet) scenarios affect four contact media each (hard floors, carpets, dust, and air). In addition, the untreated/treated concentration ratio is needed for each of the four time intervals, with the same ratio applied to all media except air. Thus, 20 input distributions are required for each indoor scenario. The outdoor scenarios affect two media each (one surface and one soil), requiring eight distributions to be specified. No untreated concentrations are modeled, so no additional ratios are needed. The pet scenarios affect only one contact medium, thus requiring just four distributions each.

3.7.4 Time Series method

If the TimeSeries option for concentrations is used, SHEDS bypasses the scenario application logic and expects the user to supply a ready-made collection of sets of concentration time series in SAS format. Each set contains a time series for each potential contact medium. The model will randomly assign a single set to each simulated individual. The person running the model determines the fraction of the population that applies pesticides lightly, moderately, heavily, or not at all based upon the relative frequency that such time series appear in the set supplied to the model. SHEDS does not currently allow the mixing of these time series with usage probabilities. A sample concentration time series data set called “timeseries” is included in the \default directory.

SHEDS is quite flexible in the time series it will accept. A “non-user time series” is easily created by simply utilizing zeroes (or background only levels) for all concentrations. One may supply concentration time series for an entire simulation period. For example, one could check the effects of light vs. moderate vs. heavy application regimes by supplying a time series of each type to the model. Alternatively, one may decide to model a specific population subgroup as represented by a set of time series sharing certain properties.

SHEDS users may create their own “timeseries” data set (outside the interface) if they wish to use this option with their own input data. This data set must contain a variable called “date,” which indicates the date in SAS date-time format. Such dates appear on the screen like “19JUL2006,” for example, but the user should be aware that these are not stored as character strings, but as integers. The other variables on the file have names consisting of a contact medium, followed by and underscore and a number. The number represents the “house” to which the measurements belong.

Suppose a field study measures concentration time series in a set of N houses. This produces N concentration time series for each contact medium. For each individual generated in SHEDS, a “house” number from 1 to N is randomly selected, and all of the concentration time series from that house are assigned to that person. This prevents the mixing of incompatible measurements, which would occur if each contact medium were assigned at random. If the user does not have multiple examples of time series for each medium, then effectively there is just one house and all the variable names (other than the date) should have the suffix “_1”.

The data for all houses appears on the same SAS data set. Each house must consist of an unbroken time series (no internal missing values) with one measurement per day. The houses do not need to have the same start and stop dates. This corresponds to the situation where the houses had unequal

(possibly staggered) measurement periods. SAS missing values (periods) are used to indicate that a particular house was not sampled on a given day. The records should be in date order. An example of a valid timeseries data set is shown in Table 2-3.

Any contact media not appearing on the timeseries file are assumed to have zero concentration. The list of media on the timeseries file may differ between houses; for example, some houses may have lawn or garden data while others do not. As currently implemented, the time series method only supports daily concentration time series and does not allow variation in concentration within each day.

3.8 Exposure-Related Inputs

The user can enter values for SHEDS input distributions in the SAS code or through the GUI (see User Guide). This section describes the different types of model inputs that relate to exposure.

3.8.1 Physical exposure factors

The potential for dermal exposure depends on the surface area of the two dermal compartments in SHEDS, which are the hands and the rest of the body. The skin surface area depends on height, weight, age, and gender. Weight is in units of kilograms (kg) and height is in centimeters (cm). Also, a basal metabolic rate and a basal inhalation rate are calculated for each individual.

The data for the calculation of these variables is on the input file "bwsa" which contains 200 sets of inputs (one set for each year of age from 0-99, for each gender). For all ages the weight is lognormally distributed, with the mean and standard deviation of log(weight) taken from the "bwsa" file. There are two approaches for height, one for children (persons up to age 20) and one for adults. For children, height is independent of weight and is normally distributed, with mean and standard deviation which depend on age and gender. For adults, a height-weight regression equation is used, with a slope, and intercept, and a normally distributed residual error. The slope, intercept, and standard deviation of the residual are taken from the "bwsa" file.

Each person is given a specific body weight sampled from a lognormal distribution, with parameters that are age and gender-specific. At present the weights are not capped at minimum or maximum values. A test was conducted, generating body weights for 10,000 adults using the SHEDS data. The minimum weight was 35 kg or about 77 pounds, which is not unreasonable. The maximum was less than 300 pounds, which if anything might be somewhat low for the current American population. Hence, there does not appear to be any need for truncating these distributions. NHANES III (http://www.cdc.gov/nchs/nhanes/nh3data.htm) was the source of the data used to develop these distributions. NHANES is designed to produce nationally representative data. Regression equations were developed for body weight, height, and surface area that preserve correlations among those parameters for a given person.

The basal metabolic rate (bmr) is calculated from a regression equation using body weight as the independent variable. The units for bmr are megajoules per day. The slope, intercept, and standard deviation of the residual are taken from the "bwsa" file. A minimum value of 0.1 megajoules per day is permitted. The basal inhalation rate is the rate in effect for activities with a METS of one and has units of cubic meters of air per hour. The basal alveolar ventilation rate, bva, is related to the basal metabolic rate by

Equation 3-1

$$\mathbf{bva = bmr * 0.166 * 0.01963 * (0.20 + 0.01* u) * 60}$$

The factor 0.166 converts from megajoules per day to kilocalories per minute. The factor 0.01963 converts from liters of oxygen consumed to cubic meters of air inhaled. The variable "u" is uniformly distributed between zero and one, and then term (0.20 + 0.01* u) represents the metabolic efficiency (liters of oxygen consumed per kilocalorie expended). The final factor of 60 converts from a per minute rate to a per hour rate.

The dermal surface area is based on the height and weight, using one formula for children below age 6 years and another for all older persons. The skin surface area (SA) is measured in square centimeters (cm^2). Below age 6 years the formula is

Equation 3-2

SA = 266.7 * (height ^0.38217) * (weight^0.53937)

For those ages 6 years and above:

Equation 3-3

SA = 305.0 * (height ^0.35129) * (weight^0.54375)

Once the total skin surface area (SA) is determined, the surface area for the hands alone is taken to be 4% of SA, while the body surface area is 96% of SA.

The input variables for the probability of having a lawn, probability of having a vegetable garden, and probability of having a pet may be considered either to be application-related inputs or as physical exposure factors. Each of these is assigned a binomial distribution in which the first parameter (called v1 in SHEDS) is the probability of a "yes" outcome, while the second parameter (called "v2") is the probability of a "no" outcome. Clearly, the sum of v1 + v2 must be one. Each individual is then tested against these probabilities, for each of the three variables.

All of the physical exposure factors are set once per person. This version of SHEDS does not age the individuals, and does not change height, weight, or other such variables over the course of the simulation period. Therefore, SHEDS version 4 is not configured for lifetime simulations, nor for multi-year simulations of children.

3.8.2 Activity-related or house-related factors

These inputs are variables that describe aspects of personal behavior or properties of the house that may influence exposure. Chemical usage and chemical transfer factors are not included here. The full list of these inputs may be found in Table 3-1 near the start of Chapter 3.

The first four variables in this category are time fractions that are used to determine contact durations. These are for carpet, lawn, vegetable garden, and pet, respectively. For example, suppose that people who have gardens spend an average of 20 minutes per day there, out of a daily average of 80 minutes of out-home time. The mean time fraction in the garden would then be 0.25. This variable should have a distribution, as some people will spend more than 25% of their out-home time in the garden, while others spend less. All four of these variables are sampled once per person in SHEDS; the results represent the mean behavior for that person. Within-person variation around this personal mean is produced using the ramp function, sampled hourly, and this maintains consistency of behavior for individuals. Note that the average amount of time in the in-home and out-home micros can be obtained directly from CHAD. Hence, given data for the mean time in each category, the time fractions may be obtained.

The maximum number of days between baths is a probability vector of length seven. "Baths" in SHEDS include showers and swimming, as these will tend to have similar effects on removing dermal loading. The first element of the vector is the probability of having at least one bath (or shower) every day. The second element is for having a maximum interval of two days, meaning that if 48 hours pass on the longitudinal activity diary without an explicit bath/shower/swim event, then a "bath" is automatically generated. The remaining elements of the probability vector cover longer intervals, up to a maximum of 7 days. This variable is sampled once per person. Since some (but not all) of the studies in CHAD explicitly report baths, the simulated individuals in SHEDS will sometimes have baths before the maximum interval has been reached.

The mean number of hand washes/day per person is considered to be a personal behavioral characteristic, and is sampled once per person. The actual number of hand washes in SHEDS changes from day to day, fluctuating around the personal mean. Actually, SHEDS only permits a maximum of one hand washing event per hour, but this distinction only matters for the upper tail of the distribution. Such persons are likely to be at the low end of the dose distribution due to low average dermal loadings, so the exact number of hand washes is not very important in those cases.

The distribution for daily soil ingestion rate should be appropriate for a 4-year-old child (see section 3.5). This distribution is sampled many times to determine the conversion factor between chemical concentration in soil and the equivalent chemical loading on the ground. The macro %SoilProperties which samples the variables in the surface-to-skin and hand-to-mouth transfer equations thousands of times, thereby determining the mean rate at which soil is transferred to the mouth. There is one unknown in this equation, namely the effective thickness of the transferred soil layer. Given the target daily soil ingestion rate, this thickness can be calculated. SHEDS uses this result to convert the chemical concentration is soil from units of μg/g to μg/cm2. The equivalent indoor conversion is handled by having the user input the dust loading on floors (separately for carpets and hard floors) directly. Such data may be available from vacuum or wipe samples in field studies. Note that the dust loading (the amount of dust on the floor) is a different variable from the chemical concentration in the dust.

There are three similar inputs, one each for the probability of having a lawn, garden, or pet. Each one is sampled from a binomial distribution. The two arguments are the probability of a "yes" outcome (v1) and the probability of a "no" outcome (v2), which must sum to one. In version 3 these were classified as "point values" instead of binomials, but as explained above (in the section on point values), "points" are reserved for variables for which all persons have the same outcome. While it is true that all persons in SHEDS have the same probability of having a pet, the outcomes may be different as some will have pets whereas others do not.

There are four similar inputs on the size (meaning the surface area) of various contact media: the house, lawn, garden, and pet. These variables are used in the handler exposure equations. The post-application exposure equations use internally calculated contact probabilities but do not directly require size variables. Hence, if the simulation only covers children (who are not handlers), or if the handler exposure switch is off, then these four inputs are not relevant to the model run. All four size variables have units of square meters (m2), unlike the concentration variables which typically use square centimeters as their unit area.

For the variable-dates option, three input variables are re quired for the distributions of the re-entry prohibition or ban time; one each for indoor, outdoor, and pet scenarios. In version 3 these were restricted to probability vectors whose return values referred to ranges of hours. The same data should not be used in version 4, which returns values that are interpreted as specific numbers of hours. For example, if a re-entry time distribution returns a value of 8, then an 8-hour minimum time until contact is allowed. The SHEDS code automatically rounds returned value down to the nearest hour, so continuous distributions may be used, although the actual ban lasts for a whole number of hours.

3.8.3 Non-chemical specific exposure factors

Another category of SHEDS inputs is non-chemical specific exposure factors. Most of the variables in this category relate to the transfer of chemical onto or off the human body. SHEDS version 4 includes a default file for non-chemical specific exposure factors (see Appendix G).

The user may choose to use either dermal transfer coefficients (TC) or transfer efficiencies (TE) to model the transfer of chemical from environmental surfaces onto the skin. Either way, the user has the option of specifying a single distribution that applies to all surfaces, or else a set of surface-specific distributions. The same TE distribution is used for both the hands and the body, although these will generally have different skin-surface contact rates (see below). When the TC method is selected, a single TC value is used to determine the total dermal transfer, but another input variable called f_DermalHands splits this amount between the hands and the body. Similarly, f_DermalHands splits the dermal exposure during handling between the hands and the body. With either the TC or TE approach, the user should be careful to determine whether the source of the data for these variables was referring to the total amount of chemical applied to the contact medium, or to the amount of chemical that was "dislodgeable" from the surface, which in some cases could be a relatively small fraction of the chemical applied.

A new output variable in version 4 is "chemfrac" which is the fraction of the chemical mass in the environment that is transferred onto the simulated person. This can be used to check that mass conservation is not violated: that the mass transferred is less than the amount in the environment. This variable should in general remain well below 1%, apart from a few exceptional cases.

Both TC and TE are usually considered to be chemical-independent because of limitations in available measurements data. However, in common with most other SHEDS inputs this default can be altered, and either TC or TE can be made chemically-dependent by adding conditions that are keyed to the chemical name.

The hand-surface dermal contact rate represents the rate at which the hands contact the relevant environmental surfaces (typically the floor or the ground). This is measured as a fraction of skin area per unit time (the units are 1/hr). Values over one are permitted and imply multiple contacts by the same patch of skin. This variable is used only with the TE approach, which implicitly measures transfer on a per-contact basis. By contrast, the TC approach uses an overall average transfer rate and the event duration to determine the amount of transfer. The body-surface contact rate serves a similar purpose, but usually the rates for the body are lower than for hands.

The hand-to-mouth transfer efficiency (HMTE) is used during hand mouthing, to determine the fraction of the loading on the mouthed portion of the hand that is transferred to the mouth (and then swallowed). The object-mouth transfer efficiency (OMTE) is used for a similar purpose during object-to-mouth contact events. Both are unitless and are sampled once per hour.

The maximum dermal loading is measured in units of micrograms of chemical per square centimeter of skin. The same loading limit is used for the hands and for the body, although the body can hold a greater mass of chemical due to its greater area. When new exposure pushes the dermal loading over the limit, the loading is immediately reduced to the limit. The limit also serves to reduce the transfer rate onto the skin. If a person is already at x% of their maximum loading, then the amount of transfer is reduced by x% from its base rate. The maximum dermal loading is sampled once per person.

The fraction of body unclothed does not include the hands. It is used to adjust the transfer rate of chemical onto the body. The assumption is that the basic transfer rate applies when 50% of the body is uncovered. When more of the body is uncovered, this basic transfer rate is increased proportionally; conversely, when less skin is uncovered the transfer rate is reduced proportionally. This roughly matches

the assumption that residential handlers wear shorts, short-sleeve shirts, socks, and shoes (see US EPA (2006b)). This variable is sampled once per day.

The hand-mouth contact area fraction represents the fraction of the surface area of one hand that enters mouth during the usual mouthing behavior for the individual. For guidance in setting this variable, the thumb has about 10% of the surface area of one hand, as does each finger. This variable is sampled once per person.

The hand-mouth contact frequency represents the number of hand-mouth insertions per hour. Each insertion is assumed to result in some chemical transfer, although SHEDS assumes that there is no replenishment or redistribution of chemical on the hand until the end of each diary event. Hence, multiple insertions during one diary event lead to diminishing amounts of transfer, as the remaining chemical loading is less with each insertion.

The object-mouth contact area represents the surface area of toys and other objects in contact with the mouth during object mouthing events. If a pliable object is crumpled or folded to get it into the mouth, then only the external surface area should be counted, since chemical removal from the internal surfaces will be minimal. For a sense of scale, note that ping-pong balls have a surface area of about 45 to 50 cm^2. The object-mouth contact area is sampled once per person.

The object-mouth contact frequency is similar to the hand-mouth contact frequency. For multiple insertions on one diary event, SHEDS assumes that the same part of the same object is repeatedly mouthed. On a new diary event, SHEDS assumes that either a new object is mouthed, or equivalently, the object has its chemical replenished by contacting the floor.

The brush-off removal efficiency can be applied to events without any washing or bathing, while awake. It allows for the gradual reduction of dermal loading over time by general activities. The fraction is expressed as a per hour rate, and is multiplied by the event duration (in hours) to find the fraction removed on each event. If the user does not want this effect included (i.e., if insufficient data are available), this variable can be set to a point value of zero. This variable is sampled once per person.

The hand-washing removal efficiency is used to reduce the hand loading on diary events with hand washing. This is not dependent on event duration. It is a unitless fraction, and is sampled once per person. The bathing removal efficiency is applied during bath/shower/swim events. It is applied to both the hand loading and the body loading. It is a unitless fraction, and is sampled once per person.

3.9 Dose Factors

The dose factors in SHEDS are used in a simple built-in first-order pharmacokinetic (PK) model to estimate the rates at which chemical moves through the body. The SHEDS PK model produces daily estimates of the amount of chemical entering the blood, and the amount of metabolite in the blood and urine. All of these variable values may be chemical dependent, in which case the use must enter conditions specifying the appropriate chemical agent associated with each distribution.

There are two dose variables associated with the dermal pathway (unlike version 3 which had only one, i.e. dermal absorption from skin to blood). The new variable, *dermal_binding*, has units of 1/hr and reflects the rate at which the dermal loading binds into the *stratum corneum*, which is the outermost layer of the skin. Once the chemical binds, it is removed from the dermal loading and it is assumed that it can no longer be washed off or transferred into the mouth. The dermal binding adjustment is made on every diary event.

The second dermal variable is *f_DermalBlood*. This is unitless and represents the fraction of dermally-bound chemical that enters the blood. All of the blood and urine estimates in SHEDS are made once per day. Thus, daily totals for the amount of dermally-bound chemical are obtained by summing over all the diary events each day, and then the chemical blood dose is obtained by multiplying this total by f_DermalBlood.

Two other variables determine the blood dose from the other pathways. *F_InhalBlood* is the fraction of the inhaled dose that enters the blood, and *f_IngestBlood* is the fraction of the ingested dose that enters the blood. Both of these operate on daily totals, like f_DermalBlood.

The variable *metab_ratio* is the molecular weight ratio of the main metabolite to the parent chemical. This ratio is specified on the *chemicals* data set, which accepts only point values (not distributions). In field studies, sometimes a metabolite is detected in blood and/or urine samples, rather than the original chemical agent. In SHEDS, the variables f_DermalBlood, f_InhalBlood, and f_IngestBlood multiply the pathway-specific doses to determine *Blood_Chemical*, which represents the the daily amount of the parent chemical absorbed into the blood, assuming no loss via metabolic transformation. This amount is then multiplied by metab_ratio to obtain *Blood_Metabolite*, the equivalent amount of metabolite, assuming full conversion. The units for Blood_Chemical and Blood_Metabolite are both ug/day. Note that the variables Blood_Chemical and Blood_Metabolite effectively double-count, since each molecule entering the blood is added to both totals (once before chemical transformation and once after). The simple SHEDS PK model does not attempt to model the metabolization processes within the body, i.e. partitioning between the parent and metabolite forms.

The final dose variable in SHEDS is *f_Urine*, which is the fraction of *Blood_Metabolite* that is excreted in urine on a daily basis. Thus, the output variable Urine_Metabolite measures the amount of metabolite excreted in urine in units of ug/day. All the blood and urine variables also have corresponding DD_ variables, such as DD_Blood_Metabolite, which convert the units to mg/kg/day. These are discussed further in the section on SHEDS output variables.

4 OUTPUTS

SHEDS saves its output as SAS data sets, with one exception (the *for_PBPK* files are saved in both SAS and ASCII text format). All permanent data sets are stored in the \output folder under the run name and run path for the job. In addition there are many temporary SAS files in the SAS Work directory that will be erased when the SAS session is closed or when another SHEDS run is performed. Since each SHEDS run name is unique, each run is written to a different folder. Caution should be used in moving these files, because the folder location may be the only indicator showing which run produced a given output file.

Each SHEDS run produces a *diarycounts* file and a *persons_demog* file. The diarycounts file enumerates each cohort-diary pool combination and the number of one-day activity diaries that are suitable for selection from that pool. The diaries are samples of behavior, and any sample may be constrained or unrepresentative if it is too small. Pools with fewer than 10 diaries are questionable in this regard. A minimum of 50 diaries per pool is preferable. Two points should be noted. First, cohorts 1 and 2 are for children below one year and are not modeled in SHEDS version 4; thus it does not matter how many diaries are in those pools. Second, SHEDS does not take any action if the pools are small. Even a pool with one diary would not prevent SHEDS from completing a run.

The persons_demog file contains a summary of all the properties assigned to each simulated person that do not change for the specified simulation. In particular it contains their age, gender, body weight, and housing characteristics among other variables as listed in the file. No exposure or dose results appear on this file. The persons_demog file is useful for subsetting the output data into various groups for analysis. For example, one might want to compare the exposure statistics of those persons who have lawns to those that do not. The person_demog file lists which persons have lawns, which allows the division of the simulated persons into these two groups. The person-numbers are common to all SHEDS output files.

All of the other permanent SHEDS output files (except the two mentioned above) are chemical-specific. The chemical is attached as a suffix to the file name. For example, a run containing chemicals X, Y,and Z would produce data sets named expo_X, expo_Y, and expo_Z, among others. These data set names are important because the variables on these data sets have generic names that do not indicate the chemical to which they apply. See the SHEDS-Residential User Guide (Isaacs et al., 2010a) for further discussion on SHEDS output and options for analyses.

4.1 Individual Time Series

Exposure time profiles for individuals are the basis of the SHEDS-Residential exposure and absorbed dose calculations. These can be viewed as plots of instantaneous exposure against time that preserve both within-day peaks and variation over longer time periods (see Figure 1-1). The SHEDS-Residential approach of tracking the sequence and pattern of new exposures and various removal processes allows for the development of a more complete picture of the source-to-dose relationship. For example, the consequences of changing the frequency or timing of hand washing and bathing events can be investigated. Also, preserving the within-day exposure variability allows for more detailed exposure inputs to PK or PBPK models.

Two output data sets contain event-level time series of exposure for all simulated persons; these are meant to be used as input to a detailed PBPK dose model. For runs with large numbers of persons, these data sets will become extremely large (each is over one megabyte per person). With more than a

few thousand persons, these files may be unmanageably large, although the limit may depend on the details of the computer system. If file size becomes a concern, SHEDS could be run several times with the same inputs but different random number seeds, and the results could be pooled later on.

The production of the event-level output files is controlled by a switch in the job settings. To suppress them, the user can set for_PBPK=0 to suppress these files and for_PBPK=1 to create them. In principle, these files may be created in uncertainty runs or multichemical runs, although these types of runs increase the amount of output. In such cases, each pass through the uncertainty loop and each chemical creates its own event-level file. For example, the file *for_PBPK_1_permethrin* contains the event-level time series of permethrin exposure for all persons, for the first pass through the uncertainty loop. The files for other chemicals and other uncertainty iterations would be indicated by appropriate changes to the file names.

The other output data sets from SHEDS provide either personal summary data (one value per variable per person), or else daily time series for each person. The only exception is the *expo_X* data set, where X is the name of the chemical agent. This file represents the full event-level output including many input variables and intermediate variables, for just the final simulated person. This file contains many more variables than the for_PBPK files, although limited to one person. If there is a person of special interest from a given SHEDS run (e.g., high end exposure value), it is possible to obtain more information on this person by re-running the same job with the same random number seed, requesting that the run stop with this person. Alternatively, one can request more data on this person using the *person_numbers* job setting option.

The *daily_X* data set is a daily summary of the expo_X data set, covering just the last simulated person. Both the expo_X and daily_X data sets are overwritten when the next person is simulated. Apart from exposure and doses in units of micrograms per day, the daily_X data set contains doses in units of mg/kg/day and margins of exposure (MOE) results. The *alldays_X* data set has the same set of variables as daily_X, but it covers all the simulated persons in a variability run. In a two-stage or uncertainty run, the alldays_X data set covers all persons in the last pass through the uncertainty loop.

4.2 Population Variability Distributions

One of the main purposes of SHEDS is to calculate the distribution of exposure across the population of interest. The model is designed so that each individual is randomly selected, but representative of the specified population. Therefore, a collection of any number of such individuals is also a representative sample. These may even be collected across multiple model runs, as long as the model input settings are the same.

The output directory contains two data sets of exposure and dose results summarized at the person level. One is *persons_mean_X* which contains the means over all simulation days (but within each person) of all time-varying daily summary variables. For example, the variable DD_Total on the *persons_mean_X* data set reports the average daily dose of chemical X in (mg/kg/day), for each person. The *persons_max_X* file is similar except that is contains the maximum single-day value for each time-varying variable, for each person. For example, DD_Total on *persons_max_X* gives the highest daily dose in (mg/kg/day), for each person. This will typically happen on or soon after a chemical usage event. However, the day in question will differ from one person to another; and furthermore, may differ from one variable to another, even within the same person. For example, the day of highest dermal exposure might or might not coincide with the day of highest inhalation exposure.

Finally, the output data set "loops_mean_X" contains the average across all persons of the exposure and dose variables, for each pass through the inner loop. Thus, this file has only one record for

a single stage (variability) run, but multiple records for a two-stage run. This file is useful in certain types of sensitivity analysis.

The first step in examining the output is to select some exposure or dose statistic of interest. For example, the ADD (average daily absorbed dose) for chemical X is represented in SHEDS by the variable Dose_total (in units of ug/day) or DD_Total (in units of mg/kg/day) on the Persons_Mean_X file. Whatever metric is selected, it is then evaluated for all the simulated individuals in the population of interest. Since the demographic characteristics of each individual are recorded, it is possible to analyze the model output at a finer level of stratification than was used to define the model run. For example, one could plot separate variability distributions for each gender or age group, even if the model run covered both genders and multiple age groups. Obviously, the price to be paid for this disaggregation is that the subgroups will have smaller numbers of individuals. As is generally the case for random samples, the variability distribution is better characterized as the number of individuals is increased.

There are two important features of stochastic models that should be noted. First, no two model runs with different random number seeds will produce the same results, even if all the input settings are identical. The "stochastic variation" is the variation in the same output statistic when the same model run is repeated. More individuals results in less stochastic variation. The second important feature is that there is less stochastic variation in the outputs near the center of the distribution than for points in the tails of the distribution. The mean or median exposure has less stochastic variation than the 90th percentile, which in turn has less stochastic variation than the 99th percentile.

For results to be interpreted correctly, the stochastic variation should be smaller than the real effects that are being estimated. For example, suppose a given statistic such as the 95th percentile of ADD is being compared across genders or across age groups. A statement that one group was higher than another is only meaningful if the difference is not likely to be due to stochastic variation. Otherwise, another pair of model runs might show no effect, or even one in the opposite direction. As a rule of thumb, the stochastic variation in the mean or median of 1000 persons might be as large as 5%, while the stochastic variation in the 90th or 95th percentiles might be closer to 10%. If the user is interested in differences of this magnitude, then larger sample sizes should be used.

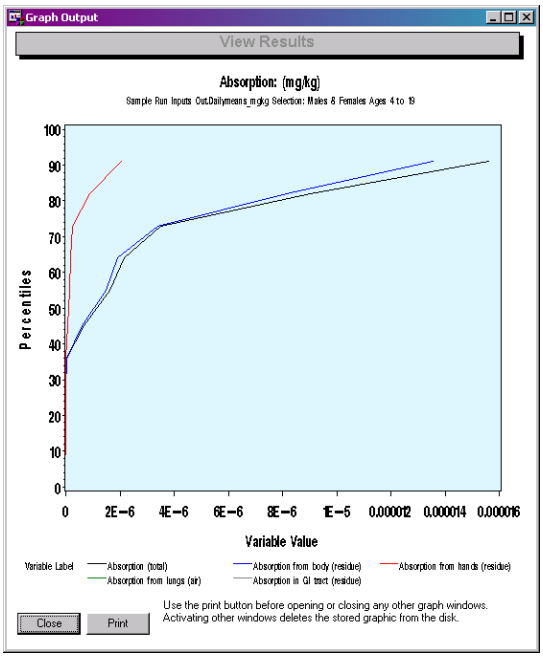


Figure 4-1. Example cumulative distribution function (CDF) plot from a SHEDS variability run.

When comparing the exposure or dose statistics across subgroups within a run, stochastic variation is always present. However, a new option in version 4 is to create pairs of model runs using the same random number seeds. This has the potential to remove most of the stochastic variation between the two groups. For example, suppose two runs have identical inputs settings except that one run is all males and the other is all females. With the same random number seeds in both runs, all the u_variables will match across runs, meaning that the same percentiles are chosen for each variable in both runs. Since most variables are independent of gender (for example, the product usage rules are in this category), this means that such variables are common to both runs. The differences in the exposure or dose distributions will be due to differences in gender, or to variables that are dependent on gender, such as body weight.

A standard SHEDS-Residential 4 variability run produces the following files in the output directory as shown in Table 4-1. An example CDF from a SHEDS variability run is given in Figure 4-1.

Table 4-1. Output files for SHEDS-Residential version 4 variability run

File	Contents
Alldays_X	One record for every person-day in a variability run, or every person-day in the last loop of an uncertainty run.
Daily_X	One record per day for just one person. If person-numbers were not specified in the job settings, this file name has no subscript and is the last person to be run. If selected Person-numbers were specified, there is one such file for each person, indicated by a numeric suffix.
Expo_X	One record for each diary event, for the same person(s) as the Daily_X file(s).
DiaryCounts	Counts of diaries available for selection in each cohort-pool combination.
For_PBPK_X	One record per diary event, for all persons. Contains only the time series needed as input to a PBPK dose model.
Loops_mean_X	Population means of daily summaries for all output variables. This file has multiple records only in a two-stage model run.
Persons_demog	One record per person with variables such as age, gender, and weight. No time-varying quantities such as exposure or dose are reported.
Persons_max_X	One record per person with maximum daily values for time-varying outputs (exposure and dose).
Persons_mean_X	One record per person with mean daily values for time-varying outputs (exposure and dose).

5 SENSITIVITY ANALYSES

One of the useful features of probabilistic models such as SHEDS is the capability to perform various sensitivity analyses. The purpose of sensitivity analysis is to determine which inputs are most responsible for the variability in the output. The inputs with the most influence on the output are the ones whose distributions should be characterized as well as possible. There is less "return on investment" in spending limited resources on improving the input distributions for variables that have little influence on the results.

The sensitivity analysis methods used in SHEDS-Residential version 4 are the same ones reviewed or recommended by the December 3-5, 2003 FIFRA Scientific Advisory Panel (FIFRA SAP, 2004), augmented by the addition of Sobol's method (Glen et al., manuscript in preparation). That SAP was requested to comment on the utility and suitability of the statistical diagnostic tools used by the SHEDS-Wood model for analyzing model results, including sensitivity analyses:

"The Panel found in general that the methods and results of the SHEDS-Wood model sensitivity and uncertainty analyses were approached in a useful and suitable manner…Since no one single method is best for examining such a complex model, the Agency's use of several different methods to examine relationships between model inputs and outputs was considered by the Panel to be appropriate because various methods can illuminate different aspects of the results. Importantly, the results of the different methods are reasonably consistent, suggesting that the conclusions of these analyses are robust with respect to choice of analytical method. Nevertheless, the results of the variability and uncertainty analyses may be limited by discrepancies between data and choice of statistical tool." (FIFRA SAP, 2004)

There are several methods of sensitivity analysis built into SHEDS. All except for stepwise regression may be run through the SHEDS graphical user interface. The correlation and regression methods may applied to either a variability or an uncertainty run. The percentile scaling and Sobol's method require two-stage model runs -- one stage for population variability and the other for selectively varying inputs. These two-stage runs generally require long model execution times, depending on number of uncertainty iterations and computing power.

5.1 The Percentile Scaling Approach

Percentile scaling requires a 2-stage SHEDS model run. The *variables* input file lists all the randomly sampled variables in SHEDS, with one record for each random variable. The user selects which ones are to be examined (controlled) by setting SensPct=1 on this file. The rest of the inputs should have SensPct=0, which allows them to vary freely. With N controlled variables, a total of (2N+1) passes through the outer loop are required; that is, (2N+1) variability distributions are modeled.

Each input variable is still assigned a distribution, whether it is to be controlled or not. The freely varying inputs are sampled in their usual fashion. For the controlled inputs, SHEDS locates three specific percentiles (the 5^{th}, 50^{th}, and 95^{th}) for each one, these being called the *low*, *medium*, and *high* settings, respectively. The first variability distribution is the *baseline* run, in which all the controlled inputs are fixed at their medium (or median) values.

For the remaining 2N variability distributions, one input is set to either its low or high setting, while all the other controlled inputs are fixed at their medians. The mean output statistic (for example, Dose_total) is calculated across all the persons in each variability distribution. Ratios of these means are

then tabulated as follows. Considering one input at a time, the ratio of mean output from the high setting to the mean output at the low setting is calculated. Similarly, the ratios of high/medium and medium/low may be calculated, using the baseline run as the medium value.

The percentile scaling approach is amenable to straightforward interpretation, due to the one-at-a-time nature of the variation of the inputs. By contrast, all the other sensitivity analysis methods allow all inputs to vary simultaneously. One issue with the percentile scaling approach is the possibility that the remaining stochastic variation from the other inputs might mask the true effect of the input in question. Another concern is that a particular input might not be very influential as long as another input is held at its median value; in some cases two or more inputs would have to be simultaneously high to have a large effect on the output. Any one-at-a-time method will miss such cases.

It is possible to run percentile scaling with a single controlled input. Then there are three variability distributions, corresponding to the low, medium, and high settings for the selected input. This isolates the variable of interest, but the remaining variables may produce enough stochastic variation to mask any effects. For example, if the high/low ratio was between 0.95 and 1.05, or even close to these values, then the true effect may be masked if the sample size was 1000 persons or fewer in each variability distribution.

By controlling more variables, the stochastic variation is reduced. The penalty for this is that the remaining variables might not provide enough variation to properly characterize the mean of the population distribution. In the extreme case that all inputs were controlled, each "variability distribution" would amount to repetitions of the exact same individual. This person might not represent the rest of the population very well. This tension between reducing stochastic variation while still retaining the representativeness of the output becomes a complicating factor when using this method. One might choose to control scenario usage, exposure, and transfer-related inputs, while allowing demographic inputs and diary assembly inputs to vary freely.

When examining the high/low ratios, note that a ratio less than one can be just as significant as a ratio over one. The latter is more common because it means that a high input value leads to a high exposure. A ratio less than one means the reverse, which is the case for processes that reduce exposure such as washing and bathing. But a variable that produces a 20% reduction in mean exposure is as influential as one that produces a 20% increase. Hence, both the high/low and the low/high ratios should be calculated (one is the reciprocal of the other), and the larger of the two should be used in ranking the importance of the inputs.

The following variables in SHEDS_Multimedia version 4 may be analyzed using the percentile scaling option:

P_lawn
P_garden
P_pet
Area_house
Area_lawn
Area_garden
Area_pet
F_carpet
F_lawn
F_garden
F_pet
Background
Dust_loading
HandWash_mean
V_bathdays
V_Reentry_in
V_Reentry_out
V_Reentry_pet
V_TimeOfUse_in
V_TimeOfUse_out
V_TimeOfUse_pet
F_Areatreated
F_UTratio
Dermal_Contact
Dermal_TC
Dermal_TE
Dermal_Maxload
Dermal_Binding
Dermal_Brushoff
F_Demalhands
F_Uncloth
F_Handmouth
F_ObjFloor
HandMouth_freq
HandMouth_TE
ObjMouth_area
ObjMouth_freq
ObjMouth_TE
F_RemovBath
F_RemovHandwash

If any of the above variables are set to point values or are not used at all (for example, dermal_TC in a transfer efficiency run), then the score will be exactly one. Variables not in the above list may be specified and will also return a (dummy) score of one, even though no analysis was performed. Do not select "mets" for analysis, as it produces a SAS error.

There are four variables related to blood and urine doses (these are f_dermalblood, f_inhalblood, f_ingestblood, and f_urine). These may be selected for percentile sensitivity analysis, but will show no effect on the usual variable for analysis, namely dose_total. These variables may show effects on other output variables.

A test of the percentile scaling method was run using the crack-crevice aerosol scenario, with variability sample size N=2000 persons, all persons are users, for a one-year simulation period. The following variables were found to be most influential on the total dose:

Table 5-1 Percentile Scaling Results for a Crack-Crevice Scenario

Rank	Variable	Low/Med	High/Med	High/Low	Score
1	Dermal_TE	0.37842	3.32139	8.77710	8.77710
2	F_AreaTreated	0.55782	1.49170	2.67414	2.67414
3	F_Uncloth	0.71598	1.38501	1.93443	1.93443
4	HandMouth_Freq	0.73729	1.20018	1.62783	1.62783
5	F_HandMouth	0.81541	1.24082	1.52173	1.52173
6	HandMouth_TE	0.79150	1.15736	1.46224	1.46224
7	Dermal_Contact	0.80033	1.16004	1.44945	1.44945
8	HandWash_Mean	1.16371	0.95166	0.81779	1.22281
9	F_Carpet	1.07412	0.92754	0.86353	1.15804
10	ObjMouth_Area	0.97432	1.09212	1.12091	1.12091
11	ObjMouth_Freq	0.97163	1.05116	1.08186	1.08186
12	F_ObjFloor	0.97104	1.02896	1.05965	1.05965
13	Dermal_MaxLoad	0.96098	1.01094	1.05198	1.05198
14	ObjMouth_TE	0.97690	1.02722	1.05151	1.05151

The column "Low/Med" gives the ratio of mean dose_total (over all persons and all days) from two runs: the Low run using the 5^{th} percentile for the named variable (and the median for the other variables in the above table) and the Med run using the 50^{th} percentile for all the above variables. Demographic variables, usage variables, and activity diary variables were allowed to vary randomly (not set to given percentiles). Similarly, the "High/Med" column gives the ratio of the mean dose_total for the High run using the 95^{th} percentile for the named variable and the median for the others, to the Med run. The "High/Low" column gives the ratio of the mean dose_total from the High run to that of the Low run. The score is either the "High/Low" value or its reciprocal, whichever is larger. The reason for this is that an x% reduction in dose is just as important as an x% increase. For example, handwashing rates are more

influential than the frequency of object mouthing, even though the former has a "lower" High/Low ratio. The higher scores indicate the more influential variables.

5.2 Pearson Correlation

The Pearson correlation is a parametric measure of the strength of the association between the settings of the input and output variables. It is calculated as:

Equation 5-1

$$\rho_p = \sum (x\,y - \bar{x}\,\bar{y}) / (\sigma_x \sigma_y)$$

where *y* is the output variable and *x* is the selected input variable. The Pearson correlation is more stable against the influence of outliers when the response variable is less skewed. Since many exposure or dose variables are positively skewed, one may optionally select to take *y* as the logarithm of a SHEDS output variable, rather than the variable itself. This often results in making the response variable closer to being normally distributed, as long as the possibility of y=0 is handled appropriately.

Usually, there is one value for *x* and one for *y*, for each person. If an input is sampled more often, then an average should be taken for each person. Similarly, a single output variable per person is required. While this is often the mean (which may be taken from the persons_mean_X file), it is possible to use the maximum daily value, or even to construct some other measure.

Pearson correlations are constrained to lie between -1 and +1. Variables that are independent of each other should have Pearson correlation of zero, although in practice such variables will report a small non-zero correlation due to stochastic or chance relationships between the samples. The algebraic sign of the correlation indicates the direction of the influence, with a positive correlation indicating that *y* moves in the same direction as *x*, and negative correlation indicating that *y* moves opposite to *x*. When ranking the influence of the various input variables, the sign of the correlation may be ignored; a correlation of either -0.5 or +0.5 is much more significant than one of either -0.1 or +0.1.

The stochastic variation in the correlation may occasionally exceed 0.05 (with either sign) for a sample size of 1000. An increase by a factor of 4 in the sample size results in halving the stochastic variation. To keep the stochastic variation consistently below 1% would require a sample size of 30,000 or more. The sample size should be considered when deciding whether one input is more influential than another, or whether the difference is not significant.

5.3 Spearman Correlation

The Spearman correlation has many similarities to the Pearson correlation; the main difference is that the Spearman is based on the association between the rankings for the input and output variables. The values of the chosen output variable are ranked across all the persons, and similarly the values of each input variable are ranked across persons (after any averaging of multiple values within each person). The Spearman correlation of the original variables is the same as the Pearson correlation of these rankings.

Like the Pearson correlation, the Spearman correlation must lie between -1 and +1, with negative values indicating that *y* moves oppositely to *x*, a zero value indicating independence, and positive values indicating that *y* and *x* tend to vary in the same direction (both up or both down together). The stochastic variation tends to be similar to that of the Pearson correlation, as long as the original variable is not strongly skewed or greatly affected by outliers. However, the Spearman correlation is less affected by the presence of outliers than is the Pearson correlation, meaning it will be subject to less stochastic variation

in that case. With Spearman correlations there is no point in taking logarithmic transforms of the output variable y since this transform preserves the rankings and hence preserves the correlation. However, if there are multiple output values for each person, transformations such as logarithms may affect the within-person averages, which could alter the rankings across persons.

The interpretation of Spearman correlations is similar to Pearson correlations. Larger correlations (in absolute value) imply greater influence. The usual procedure is to create a table listing the inputs in decreasing order of influence. Inputs with similar levels of influence may be ranked incorrectly due to stochastic variation which adds random noise to the true correlations.

Both Pearson and Spearman correlation refer to one input at a time. However, these are still considered to be multivariate methods in the sense that all inputs may vary simultaneously. Using an example from the percentile scaling approach, suppose that the response y is elevated only when two inputs x_1 and x_2 are simultaneously high. This situation could never occur under the percentile scaling approach. But the correlation methods require the inputs to be sampled independently, and on occasion both will be high simultaneously. In this example, a non-zero correlation would be seen between x_1 and y (or between x_2 and y). The correlation methods would not provide any indication that the two inputs must co-vary for y to be influenced.

5.4 Stepwise Regression

Correlation is useful for measuring the strength of the association between the output and any one given input variable. Stepwise regression also considers the combination of input variables. Stepwise regression models the output statistic as a linear function of the inputs plus a residual error. It starts by first selecting the input variable that has the largest partial R^2 correlation coefficient with the output. On the next step, another variable may be added to the model, or a variable deleted, or both. The variable added is the one with the largest partial R^2 provided that it is not already in the model. The one deleted (if any) is the one that adds an insufficient amount to the overall model R^2. This may occur if a recently added variable, despite having a reasonable partial R^2 with the output variable, has this R^2 "already included" via the correlation this input has with other inputs already in the model. The model should stop adding new variables to the stepwise regression once the partial R^2 has dropped to the level expected for stochastic variation.

Unlike correlation analysis, stepwise regression incorporates the joint effect of several variables on the output variable. The order of the variables appearing in the stepwise regression formula reflects the order of importance, with the most influential inputs appearing first. Results from stepwise regression may be compared and contrasted with the correlation methods to rank the importance of various inputs on the model's predicted dose results.

5.5 Sobol's method

Sobol's method of sensitivity analysis attempts to resolve some of the limitations of the other methods discussed above. First of all, it allows all inputs to vary simultaneously, so it can respond to cases where the output is influenced by special combinations of input values. Second, it can separate the effects of single inputs from groups of inputs varying together, which is a limitation of the correlation methods. Sobol's method can handle non-numeric inputs such as occupational category, which is beyond the capabilities of the other methods. Finally, the method does not assume or require a linear relationship between input and output. Both correlation and regression assume linearity.

Sobol's method of sensitivity analysis requires a 2-stage model run. The user defines *groups* of inputs whose collective influence is assessed. The *Variables* file has one record for each random variable in SHEDS. For each input, set *SobolGroup* to the desired group number. A group may consist of a single input, if desired. The groups should be assigned consecutive natural numbers starting with one. If there are N groups, then (2N+2) passes through the outer loop are required to determine all the main and total effects for the selected groups. All random variables must be assigned to groups. It is possible to create a "leftover" group that contains all the inputs except for the few of special interest that have their own groups. Taken to its extreme, one can create two groups, one with a single random variable and the other containing all the other random variables. This creates a valid assessment of both the main and total effects for that single input (and for all the rest of the inputs taken together).

The SHEDS code handles all the steps automatically. For each group, three output statistics are produced. For example, for group #3 these statistics are called S3, T3, and P3. The S statistics are the *main effects* for each group, the T statistics are the *total effects*, and the P statistic is a measure of the error (stochastic variation) in the other two.

Sobol's method is based on variance decomposition. Any numeric output variable which has a single value for each person may be analyzed. The variance across persons in this output variable is due entirely to the variation in the random variables, but usually some random variables are more influential than others. The total variance in the output is disaggregated into a sum over terms consisting of single groups, pairs of groups, triples, and so on. The fraction of the total variance contributed by the term representing a single group is called the *main effect* of that group. The same group also has *interaction* terms, which are multiple-group terms that include the given group. For example, if there were three groups then the total variance V would be expressed as:

$$V = V_1 + V_2 + V_3 + V_{12} + V_{13} + V_{23} + V_{123}$$

The subscripts indicate which groups are included in each term. For example, V_1, V_2, and V_3 are the main effects for groups 1, 2, and 3, respectively; and V_{13} is the interaction term for groups 1 and 3. In this example there are 3 unique pairs and just one triple. Terms with the subscripts out of order (like V_{21}) do not explicitly appear, as they are identical to terms that are present (in this case, $V_{21} = V_{12}$). The above equation is an arithmetical statement about the partial variances of the various subgroups. Similar properties are used in ANOVA and related variance decomposition techniques.

The main effect for input 1 is $S_1 = V_1/V$. The total effect for input 1 is its main effect plus all its interactions. Thus:

$$T_1 = (V_1 + V_{12} + V_{13} + V_{123}) / V$$

Since all variances are either zero or positive, all main effects and all total effects are bounded by zero and one (since they are ratios of a part to the whole). Furthermore, the total effect is equal or larger than the main effect for the same group, with equality only if all that group's interactions are zero.

Larger indices imply that the group has more power in creating exposure differences between persons. A large main effect indicates an important input, but in that case other methods (like percentile scaling) would likely show a similar result. When the main effect is small but the total effect is large, then the group is influential only through interactions with other groups, which may not be apparent with the percentile scaling or the correlation approaches. If both the main and total effects are small, then the group is not important. The definition of "large" and "small" effects depends on the number of groups. With two groups a large effect would have to be at least 0.5. But if there were dozens of groups then an index of 0.1 or even less might be relatively large compared to the others. The sum of all main effect indices can never exceed one; but conversely, the sum of all total effects indices can never be less than one.

The P statistic may be positive or negative. If the analysis were repeated many times with the same groups but different random numbers, P would be randomly distributed around a mean value of zero. Also, P would go to zero as the number of persons in the variability distribution is increased (the variance of P is inversely proportional to the number of persons). The size of P is an indicator of the possible error in the estimates of S and T. Due to stochastic error, with small numbers of persons it is even possible that the estimates for S and T may fail to be between zero and one (although their "true" values are so constrained). This can always be resolved by increasing the number of persons in the sample.

In practice, Sobol's method uses two sets of random samples from each input distribution. These may be called the "A" and "B" samples. The designations are arbitrary; that is, before the analysis one could interchange A and B and the results would be equally valid. Groups never mix their samples, so every input variable in a group uses its A sample (or else they all use their B sample) for a given pass through the model.

Each pass through the model produces one variability distribution. The essence of the method is to construct variability distributions using the A sample for some groups and the B sample for others, in a systematic way. For any pair of variability distributions, some groups of inputs will have been common to both (either the A sample or the B sample inputs were used to construct both distributions), while other groups will have used different samples. The "same" persons from both runs are compared (this means the persons with the same sequence numbers; first with first, second with second, and so on). The persons may or may not appear to be the same, as this depends on which random samples were used to produce them. Since the set of groups capture all the randomness in the model, if all inputs came from the same samples in both variability distributions, then the outputs would be exactly the same as well, and the correlation between corresponding persons in both distributions would be one. On the other hand, if different samples were used for all inputs, then the expected correlation would be zero, and any observed correlation would be due to stochastic variation and could be reduced toward zero by including more persons in each distribution.

The interesting cases are when some groups of inputs are common to both variability distributions, but other groups use different samples. If the groups using common inputs are influential, then the correlation in the output should be large (close to one), while if the groups using common inputs have little importance in determining exposure, the correlation in the output should be close to zero. The Sobol indices are derived from these correlations. By systematically varying the groups that are common to two runs, different indices may be evaluated. The code is set up to handle these steps automatically.

The correlation and regression methods can be run using a single pass (a standard variability run), although in practice these methods can also be applied to uncertainty runs that have multiple iterations. The percentile scaling method requires multiple passes (a total of 2N+1 passes for N inputs). Sobol's method requires 2N+2 passes for N groups of inputs. But in practice, the number of passes may be greater than this because the groupings may be tweaked by the user after the first results become available. The full SHEDS model has about 100 different u-variables (randomly sampled variables). In principle, each u_variable could be its own Sobol group, meaning that roughly 200 passes would be required to obtain all the main and total effect indices for all inputs. This would be a time-consuming task but is within the realm of possibility. In theory, the u_variables with multiple random samples could be further separated into groups for distinct samples, but this is not only computationally impractical but would be difficult to interpret as well, so it is not allowed in SHEDS.

Table 5-2 Sobol indices for crack-crevice aerosol scenario, for the daily mean of dose_total

Ages 1-99, variability sample size = 1500 persons, all are users of crack-crevice aerosol scenario (and no other scenarios), one year simulation period			
Variable	Main effect	Total effect	Percent
Freq_CC_Aerosol	0.124	0.539	28.6
Age	0.120	0.446	23.7
Dermal_TE	0.093	0.335	17.8
F_AreaTreated	0.020	0.111	5.9
HandMouth_freq	0.004	0.104	5.5
F_HandMouth	0.0005	0.076	4.0
DiaryT	0.010	0.051	2.7
F_Uncloth	0.009	0.044	2.3
F_Decay_In	0.005	0.040	2.1
Gender	0.0003	0.039	2.1
V_BathDays	0.008	0.033	1.8
BodyWeight	0.002	0.013	0.71
HandWash_mean	0.0008	0.013	0.70
Amount_CC_Aer	0.0011	0.0073	0.39
F_Carpet	0.0009	0.0072	0.38
Dermal_Maxload	0.0002	0.0067	0.36
Dates_CC_Aer	0.00008	0.0033	0.18
F_ObjFloor	0.00002	0.0032	0.17
ObjMouth_Freq	0.00001	0.0025	0.13
DiaryNums	0.00007	0.0020	0.11
InTreatedRoom	0.00008	0.0016	0.08
ObjMouth_TE	0.00003	0.0015	0.08
F_UTratio	0.00044	0.0007	0.04
F_RemovBath	0.00012	0.0005	0.03
Height	0.00003	0.0002	0.01
HandWashing	0.00000	0.0002	0.008
OnCarpet	0.00000	0.0001	0.005

Variables are ranked by total effect. All variables with total effect > 0.0001 are shown. The two most influential variables are the number of times per year the chemical is applied and the age of the simulated person.

The Main Effect is the fraction of the model output variance that is explained by varying the indicated variable alone. The sum of main effects over all inputs is less than one, because no interaction terms are included.

The Total Effect is the fraction of the model output variance that is explained by varying the given variable, either alone or in combination with other variables. Effects due to multiple inputs acting together are called "interaction terms". The sum of the total effects over all variables exceeds one because the interaction terms are counted multiple times. For example, if there happens to be a non-zero interaction term between Freq_CC_Aerosol and Dermal_TE, then this term would be included in the total effect on both line 1 and line 3.

The column labeled 'Percent' gives the percentage of the sum of all total effects that is represented by each row. Including variables not listed in the above table, the sum of the main effects was 0.5339, and the sum of the total effects was 1.6409, for this run. Therefore, 53.39% of the variance between persons in their daily average total dose is explained by variation in inputs acting alone. The remaining 46.61% is due to interactions. The sum of the total effects exceeds 1.4661 because some of the interactions involve three or more inputs and therefore appear more than twice in the sum of total effects over all inputs.

6 UNCERTAINTY ANALYSES

The uncertainty analysis methods used in SHEDS-Residential version 4 are the same ones reviewed or recommended by the December 3-5, 2003 SAP Panel (FIFRA SAP, 2004). That SAP was requested to comment on the utility and suitability of the statistical diagnostic tools used by the SHEDS-Wood model for analyzing model results, including uncertainty analyses. The December 3-5, 2003 SAP Panel (FIFRA SAP, 2004) felt that, "The bootstrap approach used to construct probability distributions representing uncertainty appeared to be implemented appropriately. Although alternative approaches are available for fitting uncertainty distributions from available data, using such methods is unlikely to yield an appreciable difference in the uncertainty that can be extracted directly from a given data set….Uncertainty analyses were conducted to obtain insight about the types and sources of knowledge gaps that contribute most to uncertainty about the endpoint of the modeling assessment. For this purpose, the Agency used Pearson and Spearman correlation analysis as well as stepwise linear regression to examine associations between the mean value of model inputs and output (FIFRA SAP, 2004)." While the majority of the Panel felt that the current approaches were appropriate for quantifying parameter uncertainty, they felt that, "Omitted from the uncertainty analysis is a lack of knowledge about the appropriate scenarios to include in the model and the algorithms (and corresponding data) used to simulate physical events (FIFRA SAP, 2004)." Additional research is still needed to quantify model or scenario uncertainty.

SHEDS-Residential may be run using either single stage or two-stage Monte Carlo sampling (see Figure 1-2. Monte Carlo sampling in SHEDS) The single stage is used for variability runs as discussed in prior sections. Following Cullen and Frey (1999), variability is defined as the heterogeneity of values over time, space, or different members of a population. Uncertainty is defined as the lack of knowledge about the true value of a quantity, lack of knowledge about which of several alternative model representations best describes a biological/chemical/physical mechanism of interest, or lack of knowledge about which of several alternative probability density functions should represent a quantity of interest. In terms of the effects on exposure and dose, variability refers to real differences between persons, while uncertainty refers to the consequences of the lack of knowledge about inputs. While improvements in the model itself and improvements in the input data may serve to reduce uncertainty, these should not be expected to reduce variability.

Running SHEDS in two-stage sampling mode allows the separate estimation of both variability and uncertainty. This requires that the input variables be assigned *uncertainty clouds*, as discussed below. It also requires considerable time and resources, as a two-stage run effectively generates a large family or suite of different variability distributions. For example, if 1000 individuals are used for each variability distribution and the second stage uses 500 sets of input distributions, then 500,000 simulated individuals must be generated. On a typical single personal computers, this could take a week or more of continuous operation (less time if multiple computers are used). The actual time can be shortened by running SHEDS simultaneously on several computers and pooling the results, or by cutting either of the sample sizes.

The SHEDS approach to uncertainty seeks to provide estimates of the collective effect of parameter uncertainty in all the inputs together on various statistics of the exposure distribution. It does not address other sources of uncertainty, such as assumptions about the model structure. However, some of this has been done with model-to-model comparisons (Evans et al., 2008; Young et al. 2008; Young et al. 2010 (in preparation)). In addition, model algorithm uncertainty could also be explored using different algorithm options within SHEDS, such as selecting the TC versus TE approach for dermal exposure. SHEDS version 4 does not provide information on the relative importance of each input with respect to uncertainty; that is achieved by sensitivity analysis.

A typical use of the SHEDS uncertainty results is to define ranges of exposure for given points (usually percentiles) of the variability distribution. For example, with 1000 persons run for 500 sets of inputs, the 90th percentile can be located for each variability distribution (that is, the 100th largest out of 1000). This results in 500 estimates of the 90th percentile of the population, one estimate for each pass through the uncertainty loop. If the inputs had no uncertainty at all, the 500 sets of input distributions would all be the same and the 500 estimates of the 90th percentile would differ only due to stochastic noise (these differences could be reduced by increasing the sample size beyond 1000). In practice, the spread of the 500 estimates of the 90th percentile indicate the combined effect of both the variability sample size (which is 1000 in this case) and of the uncertainty in defining distributions for the input variables. Similar remarks apply to any other percentile of the variability distribution, or to other statistical parameters such as the mean or the standard deviation.

Suppose a two-stage SHEDS run uses M persons and N sets of inputs. (In the above example, M=1000 and N=500.) Effectively, this amounts to conducting N complete SHEDS model runs of M persons each. For each of the N runs, each input variable is assigned a variability distribution selected at random from the uncertainty cloud. All M persons then sample from this single variability distribution, until the next of the N runs is started.

Without making the sample size for the variability portion extremely large (much larger than 1000), it is not possible to quantity the effects of uncertainty without including some stochastic noise. For the same reason, it is not possible to quantify variability alone without including stochastic noise. However, for uncertainty there is a second source of stochastic noise, due to the number of samples taken from the uncertainty clouds (in the above example, the 500 sets of input distributions). For a single stage variability run, this second set of stochastic variation is assumed to be minimized by selecting a central point from each uncertainty cloud, rather than sampling each cloud at random. This point should become clearer in the discussion below.

6.1 The Bootstrap Approach for Input Uncertainty Distributions

The SHEDS-Residential 4 interface does not have the built-in capability for bootstrap analysis. The user must therefore conduct this analysis outside of SHEDS, or else use pre-generated uncertainty clouds from other sources. However, if bootstrap uncertainty clouds are supplied by the user, the model can conduct uncertainty analyses with them. This can be done through the user interface or batch mode. To perform a two-stage SHEDS run, each input must be assigned an *uncertainty cloud* (Xue et al., 2006). Essentially, the uncertainty cloud reflects the confidence that the user has that the variability distribution assigned to the given input is correct.

If one were really certain that the variability distribution was right, with the correct parameter values, then the uncertainty cloud would reduce to multiple copies of this same distribution. In the more usual case, each point in the uncertainty cloud represents a potentially correct parametrization of the variability distribution for that input. The uncertainty cloud amounts to a set of variability distributions, each of which is given an equal chance of being representative for this target population. The cloud should span the range that could reasonably be assigned to this population, with a greater density of points in the more likely regions of the parameter space.

Each uncertain input variable has its own uncertainty cloud. Each point in an uncertainty cloud represent an entire distribution, not just an individual. For example, suppose body weight is lognormally distributed for persons of a given age and gender. The parameters of the lognormal distribution (which in SHEDS and the geometric mean and geometric standard deviation) are the two coordinates for each point in the cloud. Thus, two points in the cloud represents two distributions with slightly different geometric

means and geometric standard deviations. It is not known which is a better representation of the target population.

The uncertainty clouds are developed using a modification of the bootstrap method described in Frey et al. (2002) (Xue et al., 2006). Each input variable is analyzed separately. The method assumes that for a given SHEDS input variable one has data (measurements) from S original studies, where S>1. The spread or scatter among these studies provides some insight into the uncertainty for this input variable. The following steps provide a summary of the modified bootstrap method for a two-parameter distribution:

1. Fit one variability distribution (the "parent distribution"), estimating parameters v1 and v2 to all data from the original S studies combined, using the method of moments.
2. Fit a separate variability distribution (using the shape of the parent distribution) to the data from each of the original S studies using the method of moments. Examine the scatter plot of the S (v1,v2) pairs to get a sense of the scale of uncertainty.
3. Sample B data points from the parent distribution K different times (B is the bootstrap sample size); K is the number of samples of parameter pairs to be saved for uncertainty runs, typically 100 to 200; note that K is not necessarily the same as N.
4. For each of those K sets of B data points, fit the parent distribution and compute the parameter values. This gives K (v1,v2) pairs.
5. Overlay the scatter plot of these K (v1,v2) pairs with the S (v1,v2) pairs from step (2).
6. Repeat steps 3-5 with different values of B, until the scatter plot from step (4) satisfactorily matches the spread seen in the scatter plot from step (2).
7. Save the K sets of (v1,v2) pairs from the final repetition of step 6.

At the start of each uncertainty simulation, one of the K parameter pairs is randomly selected for each input variable. The selected (v1,v2) pairs define the variability distributions to be used for this simulation. All of the individuals within one uncertainty simulation randomly draw values from these variability distributions. This approach avoids potential problems arising from determining v1 and v2 independently.

Steps 5 and 6 above require one to compare the input variable uncertainty with that seen in the original studies, leading to a choice for B. Typically, it was found that B= 5 was suitable for very small or highly uncertain datasets; B=10 for somewhat larger datasets; and B=15 or B=20 for variables with fairly little uncertainty. The size of B depends on the sample size of the original data and the how well the bootstrap scatter plots fit over the data points among the various data sources (step 5).

The above applies to those inputs having two parameter distributions. All of the forms currently supported in SHEDS are two-parameter, except for the point (1-parameter) and triangular (3-parameter). Point values may have uncertainty (although they have no variability), as different surveys or studies may provide different estimates. Furthermore, some SHEDS runs may be applied to subpopulations that differ from the survey population. Uncertainty can and should be assigned to the one and three parameter distributions in SHEDS, as well. The bootstrap method could be used, except that the comparison of scatter plots is simplest and most natural in two dimensions. Obtaining uncertainty clouds for points is straightforward (and simpler than the bootstrap method) as long as there are multiple sources of estimates. For the triangle distribution, the bootstrap method could still be used without alteration, if the user required that one point of the triangle be fixed (for example, the minimum could be always kept at zero).

A three-parameter variant of the bootstrap is fine in principle, but it has the technical problem of plotting and matching clouds of data points in a three-dimensional space.

Each of the K points in the uncertainty cloud represents a fully-specified variability distribution. For a two-parameter distribution, each of these K points consists of a (v1,v2) pair that are to be used together.

6.2 Uncertainty inputs and outputs

For each of the N sets of inputs in a two-stage SHEDS run, one of the K points in the uncertainty cloud is selected at random, for each input. There is no need for the value of K to be the same for all inputs. In fact, if a certain input had no uncertainty, or were to be excluded from the uncertainty analysis for some reason, one could set K=1 for that input, so that the same variability distribution was always used.

Since one of the K points is selected at random, there is a chance that the same point in the uncertainty cloud will be selected for more than one of the N uncertainty repetitions. If K<N, then this becomes certain. However, this is not a concern since the other input variables are likely to be assigned different points from their clouds. Suppose that the model has 20 input variables, each with K=10 for their uncertainty clouds. Then there are 10^{20} possible combinations of input distributions, while only N of these (typically N is a few hundred) are actually sampled. In practice, there is virtually no chance that any two of the N sets of input distributions will be identical. If this happened anyway, then their respective outputs would differ only by stochastic variation, not by uncertainty.

For the 2-stage runs, the time series results for specific individuals are not retained. First of all, it would not be practical to save all the time series for MxN (perhaps 500,000) individuals. Furthermore, there is no clear correspondence between individuals on different uncertainty iterations, since the demographic variables are re-sampled for each of the N input sets. Since corresponding individuals cannot be compared, it is sufficient to compare population distributions. For this purpose, each person may be summarized by some selected statistical parameter. For example, the average daily absorbed dose (ADD) may be used for this purpose. If desired, the mean value for selected inputs can be recorded for each person, which permits a combination uncertainty/sensitivity analysis to be performed.

Without performing sensitivity analysis, there are two commonly used types of graphical analysis for the uncertainty output (Xue et al., 2006). One involves displaying the results of three of the N variability distributions, specifically those corresponding to the 5^{th}, 50^{th}, and 95^{th} percentiles of the N cases, as ranked by their medians (Figure 6-1). For example, with N=500 the variability distributions that had the 25^{th}, 250^{th}, and 475^{th} highest medians would be plotted. Each of these three curves contains the data from all M persons in that simulation. The horizontal axis represents percentiles of the population variability. The vertical distances between the three curves represent uncertainty in each percentile of the variability distribution.

The other type of uncertainty graph displays three selected variability percentiles (the 5^{th}, 50^{th}, and 95^{th}) from the output corresponding to each of the N input sets (Figure 6-2). The N values for each of the 3 variability percentiles are sorted separately from lowest to highest. Here the horizontal axis represents percentiles of the uncertainty distribution, while the vertical separation between the curves measures variability.

In SHEDS version 4, uncertainty in inputs is specified much like variables sampled from empirical distributions. Each input that has uncertainty has a pointer to a file in the \input\unc directory. This file contains one set of distributional parameters on each record, and may have any number of

records. In fact, this file is simply an empirical distribution of the points in the uncertainty cloud. The usual variables on this file are called "v1" and "v2", which are the parameter names for 2-parameter distributions. The modified bootstrap approach described above provides a systematic way to obtain these values. However, in principle any other parameters may be modified, including truncation points or the form of the distribution itself. Even probability vectors or conditions may be altered. All variables supplied on the uncertainty data set replace their corresponding values on the usual input data sets.

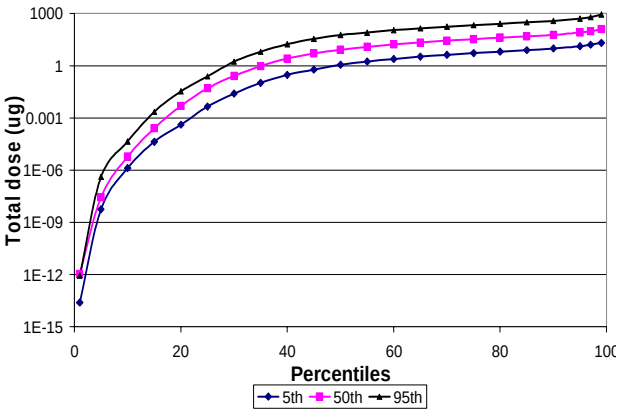


Figure 6-1. Hypothetical uncertainty analysis results.

CDFs for low-, medium-, and high-dose populations, as ranked by their medians. The horizontal axis represents percentiles of the population variability. The vertical distances between the three curves represent uncertainty in each percentile of the variability distribution.

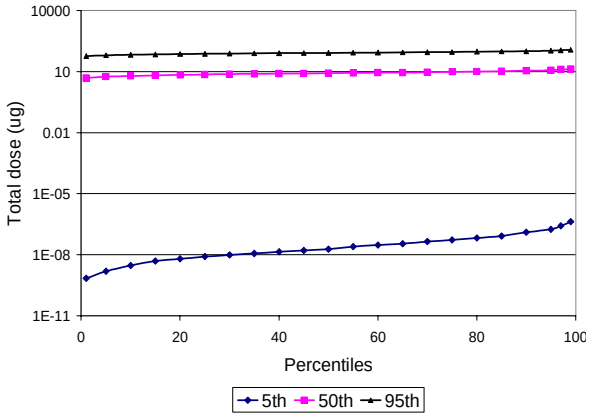


Figure 6-2. Hypothetical uncertainty analysis results.

CDFs for the 5th, 50th, and 95th percentiles from all simulated populations. The horizontal axis represents percentiles of the uncertainty distribution, while the vertical separation between the curves measures variability.

6.3 Sensitivity Analysis on Results from a Two-Stage Monte Carlo Model Run

The Pearson correlation, Spearman correlation, and stepwise regression methods of sensitivity analysis discussed in the Sensitivity Analyses section above may also be applied to the results from a two-stage SHEDS run. For this purpose, the mean value for each input variable and the mean output statistic are recorded for each of the (MxN) persons in the uncertainty run. The above methods are then applied, just as if the entire set of (MxN) persons belonged to a single-stage run.

This type of sensitivity analysis reflects the overall importance of the combined uncertainty/variability variation in each input on the output statistic. If a given input variable exhibits little importance in this regard, then the model will not benefit much from improvements in the characterization of the distribution for that input. For variables identified as important, the next step is to ascertain whether the variability or the uncertainty is predominant. This can be achieved by performing a sensitivity analysis on a single stage run. If the uncertainty is large, then the SHEDS exposure estimates can be improved by refining the knowledge of this input variable. If variability (rather than uncertainty) predominates, then this variable has an important influence on exposure, but this variation mostly reflects real differences among the target population and will not be much reduced by conducting further research into the input distributions.

7 QUALITY ASSURANCE (QA)

The development of the SHEDS-Residential, version 4 model has been governed by a Category I Quality Assurance Project Plan (QAPP) prepared for this version of the model (US EPA, 2010). The QAPP describes the overall organization and management of the model development process, background on the model and its structure, and approaches employed for the QA process. Much of this material is discussed much more extensively earlier in this technical manual. This section describes in detail several specific QA activities for version 4.

SHEDS-Residential version 4 was conceptually developed (model approach and algorithms), designed (GUI), and quality assured by exposure modeling researchers in EPA/ORD/NERL. Writing the SAS computer code to implement the model, developing some of the model algorithms, and constructing the GUI was conducted by Alion Science and Technology under Contract EP-D-05-065. Integrated throughout this entire process was a heavy emphasis on QA, both by EPA and Alion Science and Technology. This emphasis resulted in a high degree of reliability in both model operation and GUI performance.

Apart from the ongoing QA during the development phase, four kinds of quality assurance tests were run on SHEDS Multimedia version 4. The first kind of tests involve a systematic check of the various modeling options. A small number of persons are run for each test and the SAS log is checked for warnings and errors, and the output files are checked for missing or unreasonable values. With only a few persons, the output numbers cannot be reliably compared to those obtained with other model versions since the stochastic variation is too large.

The second kind of test involves running many persons through the same scenario in both SHEDS version 3 and version 4. A crack&crevice aerosol application of permethrin pesticide was chosen for this test. Two thousand individuals were simulated, which was sufficient to reduce the stochastic variation to a level where real differences in the model output can be discerned. Any differences were then examined and explained in section 8.3

The third kind of test involves the detailed examination of a few events for a single person, showing the steps in calculating contact, transfer, exposure, and dose.

The fourth kind of test involves running the model through the GUI, and comparing those results against those using the SAS code. This has two components: checking that the input files reflect the choices made by the user when defining the run, and checking that the output tabulation and graphing features are operating correctly.

7.1 Development QA

As the model structure was being developed, the equations were checked both individually and collectively for consistency and to ensure that they were combined properly so that the model itself behaved as intended. The model has been implemented in SAS and requires only the BASE, STAT, and GRAPH modules. Development of the GUI also required the AF module, though this piece of software is not needed to run the model or GUI. SAS software is universally recognized as an extremely reliable statistical programming language and is thoroughly quality assured before its release. In addition, SAS provides technical support to its users, and the SHEDS-Residential developers availed themselves of this utility during model code and GUI construction.

As the model code and GUI were being built, the developers cross-checked each other's work and as problems were revealed they were corrected. This checking procedure was done on individual modules and sections of code (e. g., diary assembly, random distribution sampling, exposure module), and as portions of new code were added to existing programming to make certain that the assembled unit performed correctly.

Next, individuals in ORD/NERL, OPP, and Alion tested SHEDS-Residential, including evaluating the user-friendliness and layout of the GUI. They recorded any problems or suggestions for improvement and communicated these to the developers. The model code and GUI were then appropriately revised.

7.2 Check of modeling options

SHEDS allows the user to customize a model run in many ways by altering the *job settings*. These settings are recorded on the RunInfo file, except for the chemicals and scenarios to be run which are on other input files. A series of QA tests were run with revision 4_4_3, with names starting with QA followed by a 2-digit number and an optional alphabetic suffix. Test QA0 is a base case, with all the other tests being variations from that. The number in the test name indicates which setting is being varied from the base case. The full list of job settings for the base test QA0 is the following:

Job Settings for run QA0

1	Persons:	5
2	Age groups:	All (except below 1 year)
3	Gender:	Both
4	Uncertainty loops:	None
5	# Chemicals:	1
6	# Scenarios:	10
7	Debug level:	0
8	Product usage method:	Fixed dates
9	Concentration method:	Decay-dispersion
10	Dermal transfer method:	Transfer efficiency (single distribution)
11	Diary assembly method:	Eight diaries
12	Simulation period:	Jan 1, 2010 – Dec 31, 2010
13	Random number seed:	12345
14	Model handler exposure:	No
15	Event splitting fraction:	0.5
16	Max. # prior usage days:	30
17	Percentile sensitivity:	No
18	Sobol sensitivity:	No
19	Correlation of inputs:	No
20	Save intermediate vars:	No
21	Pre-process activity diaries:	No
22	Output for dose model:	No
23	Progress indicator on log:	Yes
24	Save event-level output:	No
25	Save log file:	No
26	Save selected persons:	No

QA0 takes about 19 seconds of execution time, although this may depend strongly on the computer. The mean values for the dose_total variable over the year-long simulation are:

Person	#1	3.069 ug
	#2	10.657 ug
	#3	25.366 ug
	#4	2.255 ug
	#5	22.084 ug

The above values are based partly on artificial input distributions designed for testing purposes, and should not be taken as typical of real exposures for this chemical.

Each of the remaining runs in this set varies just one setting away from those in QA0. QA01a sets the number of persons to one. The result is a mean dose_total=3.069, as expected. QA01b requests 20 persons. The dose_total results for the first five of these 20 persons match the results from QA0. This is expected since the random number seeds for the two jobs are the same.

QA02a and QA2b examine variation in the selected age groups. QA02a is adults only (ages 20+), while QA02b is children only (ages 1-12). These restrictions are confirmed by examining the persons_demog output file.

QA03a restricts the population to females only. QA03b is males only. These restrictions are confirmed on the person_demog output file.

QA04 is an uncertainty run. Only two inputs are assigned uncertainty distributions. At the end of the model run, SHEDS creates two plots, one with variability percentiles on the horizontal axis and the other with uncertainty percentiles on the horizontal axis. The code used for producing these graphs requires at least 20 persons and 3 uncertainty loops, so these numbers are used in this test.

QA05 is a multichemical run. Three chemicals are used, the third being permethrin. The dose_total for permethrin is the same for each person as in run QA0.

The QA06 family of tests examines different combinations of scenarios. QA06a has two lawn scenarios, QA06b has one vegetable garden scenario, QA06c has five indoor scenarios, and QA06d has two pet scenarios. These runs all completed successfully. The doses cannot be directly compared to those in QA0 because the latter does not separate the doses from each scenario.

QA07 runs the code with the option debug=1. This has no effect on the calculations, but greatly increases the number of messages written to the SAS log. In particular, the *notes* and *mprint* options are turned on, which means that SAS writes summary information after every data step and procedure, and the active code is printed after each round of macro resolution. To keep the log to a reasonable length, the number of persons was reduced to two. Their mean dose_total values were 3.069 and 10.657, as expected.

QA08 replaces the Fixed Dates method with the Model-Determined Dates method. This changes the chemical usage patterns, which has a large effect on the exposures of individuals. The mean values for the dose_total variable over the year-long simulation are:

Person	#1	2.941 ug
	#2	8.596 ug
	#3	9.211 ug
	#4	0.811 ug
	#5	15.170 ug

QA09a uses the Intervals method (rather than Decay-dispersion) for determining the concentration time series. The usage dates are still fixed and are the same as in the previous tests, but the post-application concentrations are sampled from distributions that depend on the time since application (binned into day of usage, 1-7 days after usage, 8-30 days after, and 31 or more days after). QA09b uses Intervals, but with Variable usage dates (which was examined in QA08 with the Decay-dispersion option). QA09c has no usage dates and uses the Timeseries option for concentration, where the user supplies concentration time series for the relevant contact media. Note that in this example the Timeseries option is very fast (two seconds for five persons) because the simulation period is reduced to the length of the concentration time series, which in this case is only a few days.

The QA10 tests examine options for determining chemical transfer onto the simulated individual. QA10a uses the transfer coefficient approach. QA10b uses transfer efficiency, but uses separate distributions for each surface contact medium (there are five: indoor hard surfaces, indoor textured surfaces (carpet), lawn, vegetable garden, and pets). In both cases, the exposures and doses differ from those in QA0 since the transfer rates differ.

The QA11 tests use the new longitudinal diary assembly method. QA11a has time outside as the key variable and sets the diversity parameter D=0.3 and autocorrelation A=0. QA11b changes the key variable to travel time and uses D=0.1 and A=0.1. In both cases the mean dose_total variables are typically within 10% of those in QA0. Note that the diaries determine the amount of time spent near the chemical, but do not affect the chemical usage dates, the amount applied, the concentrations, or the transfer rates. Hence the overall effect of altering the diaries is not expected to be very large.

QA12a reduces the simulation period to its minimum, which is one day. QA12b extends the simulation period to cover more than one year. While the simulation period may be potentially any length, note that SHEDS does not age the simulated individuals, which could be a concern when assessing the validity of multi-year simulations.

QA13a sets the random number seed to a different value, which should create individuals that have no relationship at all to those in QA0. In QA13a, person #2 had the lowest dose_total and person #4 had the highest. QA13b sets the seed to zero, which requests the program to use the computer clock to set the seed. This clock-generated seed is written back onto the RunInfo file, to allow the user to repeat this run later, if desired. Such seeds are usually 9 or 10 digit numbers. In QA13b, person #1 had the lowest dose_total and person #3 had the highest. With such small numbers of persons the correlation (or lack thereof) in the results cannot really be determined, but there is no reason to think that the QA13a or QA13b results are tracking those of QA0.

The QA14 runs include the exposures incurred during the handling or application of the chemical, provided the simulated person is the handler. With fixed usage dates, all persons who are at least the minimum handler age are deemed to be handlers. QA14a uses fixed dates with a minimum handler age of 16, which results in four handlers out of these five persons. QA14b raises the minimum age to 30 which

eliminates the handler exposure for two of the five persons. With the Model-Determined dates option, handler status is randomly determined for persons over the minimum age. QA14c uses the Model-Determined dates option with handlers included. In QA14a and QA14b, the non-handlers have the same dose_totals as in QA0, whereas the handlers have elevated values.

QA15a sets the event-splitting fraction (also called ramp size) to zero. This means that each diary event has contact with one surface at most. The default setting of 0.5 is the most consistent setting for compatibility with SHEDS version 3. QA15b examines the other extreme of setting the fraction to one. This splits as many diary events as possible between multiple surfaces, subject to the constraints on contact probabilities. Note that the event-splitting fraction has no effect on the mean amount of time in contact with any surface, it only changes the sequence of contacts. In QA15a the dose_total means are: 3.065, 10.633, 25.343, 2.252, and 22.071. In QA15b these are: 3.063, 10.627, 25.372, 2.251, and 22.108. These compare to the baseline run QA0 with the default of 0.5, which gave 3.069, 10.657, 25.366, 2.255, and 22.086. In all cases the changes were well below 1%. The effects of changing this setting are minor.

The QA16 runs examine the effect of assigning usage dates prior to the start of the simulation period. This allows chemical to be present at the start of the simulation. For example, an application one or two weeks before the simulation would create substantial potential for additional exposure during the simulation period. This option is only relevant with Model-Determined dates, since with Fixed dates the user specifically sets all usage dates (including any that happen before the simulation period). QA16a uses Model-Determined dates with no prior usage dates allowed. QA16b allows usage up to 90 days before the simulation period begins. The mean doses over the entire year are only slightly (under 3%) affected by this, but the doses near the start of the simulation are much higher in the latter run. For example, the first person in QA16a has zero dose until the first usage event in the simulation period, which happens more than two months after the simulation starts. In QA16b this person has exposure and dose starting on the first day.

QA17 performs a percentile scaling sensitivity analysis. This requires (2N+1) passes through the outer code loop, where N is the number of input variables being analyzed. In this case, as in most of the other QA tests, the inner code loop consists of five persons. Not all input variables need to be analyzed; the user selects them by setting SensPct=1 on the Variables input file. In QA17 only two inputs were selected. SHEDS prints a summary table of the ratios of the mean dose_total at the various input settings. Ratios near one indicate inputs of minor effect, while higher (or lower) ratios indicate a larger change in output when that input is altered. A table showing the effects of the selected variables was printed.

QA18 performs a Sobol sensitivity analysis. The user indicates the input variable groups on the Variables file by assigning a group number to each random variable (each record of the file). For this test only three groups were used. The number of passes through the outer loop is (2N+2), which is eight in this case. SHEDS prints the S, T, and P indices for each group, where S is the main effect, T is the total effect, and P is the spurious correlation. These indices were printed for the three selected groups.

QA19 uses the option of correlating the randomly-sampled variables. The pairwise correlations are listed on the *correlations* input file; pairs not appearing here are assumed to have zero correlation. The NORTA method (described in the methods section) is used to generate the correlation. In QA19 the HandWash_Mean and f_RemovBath variables were correlated at a level of 0.8. The dose_total values were within about 2% of those in QA0.

QA20 saves more variables for subsequent analysis. For example, the expo_permethrin output file contains event-level data for the last simulated person (this file is overwritten when going to the next person). The standard run QA0 uses 12 MB for expo_permethrin, whereas QA20 requires nearly 30 MB for this file.

QA21 performs pre-processing on the activity diaries before starting the simulation. The diaries are already pre-processed in the default data directory, so this option is only needed if the diary database changes or if the user changes the selection/rejection rules for using the diaries in SHEDS. Running the diary pre-processing requires less than one minute of extra time and this time is independent of the number of persons being simulated or the length of the simulation period.

QA22 performs the same calculations as the base run, but it produces three extra output files (for sensitivity or uncertainty runs, there are three sets of extra files). These files are intended as input to a PBPK dose model. One file is a SAS data set containing the important exposure variables at the event level for all simulated persons. The second file contains similar information but in ASCII format. The third file is an ASCII version of important variables from the persons_demog file.

QA23 turns off the progress indicator for each person. Only the starting of the outer loop is recorded. This option may be needed for extremely large runs, when the messages for each person might fill up the SAS log.

QA24 creates a SAS data set that contains all event-level data for all simulated persons. This can easily become a very large file. For a one-year simulation the file is nearly 5 MB per person, or 1 gigabyte for each 220 persons. This may place limits on the number of persons that can be run under this option.

QA25 writes the SAS log to the ASCII file "saslog.txt" in the run output directory. No SAS log information appears on the screen. Note that this file is not erased at the start of each run, so logs of multiple runs may appear on the same file, if multiple runs are performed using the same name. In such cases the SAS data sets from the earlier runs are erased.

QA26 runs only selected persons from an earlier run. This is useful if the user wants more information on one or more selected persons. In this test case, persons 2, 4, and 9 were requested from a run of 10. Persons 2 and 4 matched those from runs QA0 and QA01b; person 9 matched the same one from QA01b (this person did not exist in QA0 as only 5 persons were modeled). To obtain the same results as an earlier run, it is necessary to copy all job settings and input files, and especially the random number seed.

The above 43 tests attempt to cover each option that may be selected when running SHEDS. However, there are many combinations of these options that could be run, and it is not practical to try to test them all. It is always possible that certain untested combinations could result in problems; when these are found, they are corrected. In some cases, combinations of options are forbidden; for example, correlation is not allowed with sensitivity analysis.

There is no limit in SHEDS on the number of persons simulated, but in practice the computer will impose its own limits. When this number becomes too large, the SAS log may fill up, the memory requirements may become excessive, various SAS data sets may become too large to handle, and the overall performance may slow down and ultimately fail. However, the baseline run can be expanded to many tens of thousands of persons before this becomes a risk. Such runs could take many days to complete.

7.3 Comparison of SHEDS version 4 with version 3

A comparison of SHEDS version 4 outputs to version 3 was run for a crack-crevice aerosol application, with 2000 persons (all 3-5 year old children with at least one application per house), simulated for one year each. The version 3 and version 4 inputs were matched as closely as possible. Table 7-1 shows the corresponding outputs in the two versions. The version 3 run used the code

obtainable as a download from the EPA SHEDS web site (http://www.epa.gov/heasd/products/sheds_multimedia/sheds_mm.html).

The comparison is complicated by changes made in the definition of the exposure-dose boundaries. In version 3, *absorption* meant *into the blood.* Chemical that had not yet entered the blood (such as chemical in the GI tract) was still *exposure*. Version 3 distinguished new exposure from running exposure (the accumulated amount still present). Version 3 also distinguished residue from matter until the chemical was in the blood.

In version 4 the exposure-dose boundary is at the external surface of the body. As soon as the chemical is inhaled or ingested it becomes *dose*. This is very different from version 3; for this Version 4 run only 36% of the ingested dose was absorbed into the blood. Also, version 4 calculates blood dose only at the daily level, not the event level.

Table 7-1. Correspondences between certain exposure and dose variables in Versions 3 and 4

V3 variable	V3 label	V4 variable	V4 label
expHr+expHm	New exposure on hands	Expo_hands	New exposure on hands
expBr+expBm	New exposure on body	Expo_body	New exposure on body
expGr+expGm	New GI tract exposure	Dose_ingest	Ingested dose
expLA	New exposure in lungs	Dose_inhal	Inhalation dose
absBr+absBm+absHr+absHm	Absorption from hands+body	abs_dermal	Blood dose
absGr+absGm	Absorption from GI tract	abs_ingest	Blood dose from GI tract
absLA	Absorption from lungs	abs_inhal	Blood dose from lungs
abstot	Total absorption	Blood_Perm.	Amount entering blood

Dermal exposures were quite close (within a few percent) in the two versions, as were the GI tract exposures. The version 4 inhalation exposure was about 20% larger than in version 3, due to a change in modeling the air in the untreated part of the house (see section 2.4.4.1 for details). Since inhalation is the pathway with the least exposure in this example, the overall change in exposure from version 3 to version 4 was quite small (a few percent). See Figure 7-1 for a comparison of the CDFs for total dose.

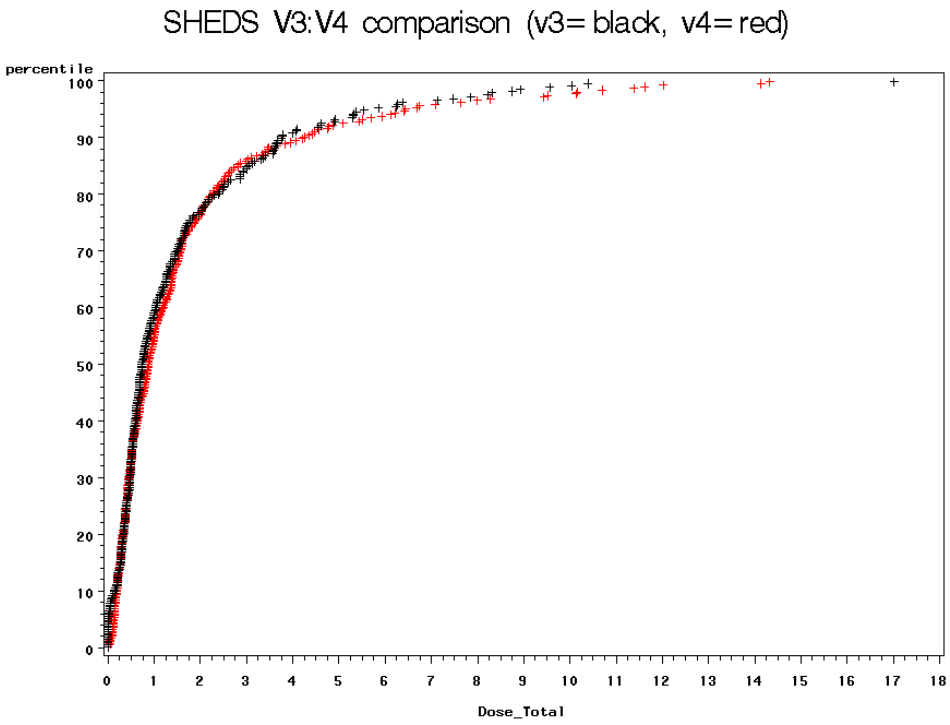


Figure 7-1. Comparison of the CDFs for total dose in SHEDS Version 3 versus Version 4.

7.4 One-person simulation for code verification

After all these steps, the implementation of the exposure equations in the final code was numerically evaluated by comparing the output from a short simulation with selected hand calculations. A model run of SHEDS-Residential was made for just one person for a partial day. This individual was a 3-year old girl. An artificial diary with 5 events was constructed, with enough variety to provide examples of nearly all of the exposure equations.

The data set containing the event-level exposure variables was examined in detail. The hand calculations verified that the exposure equations were implemented as intended.

This process for a specific diary event is illustrated by the calculations shown below. Each number was taken from the corresponding variable on the expo_permethrin data set resulting from this exercise.

Diary event 8, start=7:00, duration=20 minutes.

Dose_Inhal = conc_air * bva * mets * (duration)

0.19644 = 3.32085 * 0.16228 * 1.09351 * (20/60)

Expo_hands = (area_hands * HCR * TE) * conc_hard * (duration)* f_load

13.9513 = (279.589 * 2.58063 * 0.0302591) * 1.91706 * (20/60) * 1.0

Expo_body = (area_body * BCR * TE) * conc_hard * (duration)* f_uncloth

46.2675 = (6710.14 * 0.66169 * 0.0302591) * 1.91706 * (20/60) * 053891

Dose_hands = Expo_hands * dermal_binding * (duration)

0.0023252 = 13.9513 * 0.0005 * (20/60)

Dose_body = Expo_body * dermal_binding * (duration)

0.007711 = 46.2675 * 0.0005 * (20/60)

Dose_HTM = (RunExp_H/2) * f_HandMouth * (1-(1-HMTE)**(HandMouth_freq*(duration)))

0.07391 = ((13.9513-0.0023252)/2) * 0.13909 * (1-(1-0.12607)**(1.76413*(20/60)))

Expo_HTM = Dose_HTM / HMTE

0.58623 = 0.07391 / 0.12607

Dose_OTM = conc_hard * f_ObjFloor * OM_area * (1-(1-OMTE)**(ObjMouth_freq*(duration)))

0.14066 = 1.91706 * 0.18047 * 9.00326 * (1-(1-0.36401)**(0.30631*(20/60)))

Expo_OTM = Dose_OTM / OMTE

0.38642 = 0.14066 / 0.36401

Dose_Ingest = Dose_HTM + Dose_OTM

0.21457 = 0.07391 + 0.14066

Remove_H_bath = RunExp_H * f_RemoveBath

12.5278 = (13.9513-0.0023252-0.07391) * 0.90290

RunExp_Hands = Expo_hands – Dose_hands – Dose_HTM – Remove_H_Bath

1.3473 = 13.9513 - 0.0023252 - 0.07391 -12.5278

Remove_B_Bath = RunExp_B * f_RemovBath

41.768 = (46.2675 -0.007711) * 0.90290

RunExp_Body = Expo_body – Dose_body – Remove_B_Bath

4.4919 = 46.2675 -0.007711- 41.768

8 REFERENCES

Adgate J.L., Weisel C., Wang Y., Rhoads G.G., and Lioy P.J. 1995. Lead in House Dust: Relationships between Exposure Metrics. *Environmental Research* 70: 134-147.

American Veterinary Medical Association. 2007. U.S. Pet Ownership and Demographics Sourcebook. http://www.avma.org/reference/marketstats/sourcebook.asp

Brinkman S., Gialamas A., Jones L., Edwards P., and Maynard E. 1999. Child Activity Patterns for Environmental Exposure Assessment in the Home, National Environmental Health Forum Monographs. Exposure Series No. 1 Department of Human Services, Adelaide, Australia. http://www.publications.health.sa.gov.au/envh/50/

Burke J. and Vedantham R. 2009. Stochastic Human Exposure and Dose Simulation for Particulate Matter (SHEDS-PM) Version 3.5 User Guide, September 2009, US EPA.

Cohen Hubal E.A., Sheldon L.S., Burke J.M., McCurdy T.R., Berry M.R., Rigas M.L., Zartarian V.G. and Freeman N.C.G. 2000. Children's Exposure Assessment: A Review of Factors Influencing Children's Exposure, and the Data Available to Characterize and Assess that Exposure. *Environmental Health Perspectives* 108(6):475-486.

Cohen Hubal E.A., Suggs J.C., Nishioka M.G., and Ivancic W.A. 2005. Characterizing residue transfer efficiencies using a fluorescent imaging technique. *Journal of Exposure Analysis and Environmental Epidemiology* 15: 261–270.

Cullen A.C. and Frey H.C. 1999. Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs. Plenum Press. New York, NY.

Evans, J., V. Zartarian, J. Xue, B. Young, J. Driver, M. Pandian, N. Tulve, L. Smith, and G. Glen. 2008. Residential Exposure Model Algorithms: Comparisons by Exposure Pathway across Four Models. Paper presented at the American Chemical Society meeting, August, 2008, Philadelphia, PA.

FIFRA SAP 2004. Meeting Minutes No. 2003-04, FIFRA Scientific Advisory Panel Meeting, December 3-5, 2003, held at the Sheraton Crystal City Hotel Arlington, Virginia, "A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding: Draft Preliminary Probabilistic Exposure and Risk Assessment For Children Who Contact CCA-Treated Wood On Playsets And Decks And CCA-Containing Soil Around These Structures." http://www.epa.gov/scipoly/sap/meetings/2003/dec3/dec35meetingminutes.pdf

FIFRA SAP 2007. A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding: Review of EPA/ORD/NERL's SHEDS-Residential Model Aggregate version 3, SAP Minutes No. 2007-06. August 14-15, 2007 FIFRA Scientific Advisory Panel Meeting, Arlington, VA.

Freeman N.C.G., Jimenez M., Reed K.J., Gurunathan S., Edwards R.D., Roy A., Adgate J.L., Pellizzari E.D., Quackenboss J., Sexton K., and Lioy P.J. 2001. Quantitative Analysis of Children's Microactivity Patterns: The Minnesota Children's Pesticide Exposure Study. *Journal of Exposure Analysis and Environmental Epidemiology* 11: 501-509.

Frey H.C., Zheng J., Zhao Y., Li S. and Zhu Y. 2002. Technical Documentation of the AuvTool Software Tool for Analysis of Variability and Uncertainty. Unpublished contract report submitted to US EPA, Feb 2002 (PR-RT-02-00381).

Glen G., Smith L., Stallings C., Isaacs K., Zartarian V., and Xue J. 2007. Planned Methodologies for Extending SHEDS-Residential version 3 (aggregate only) to SHEDS-Residential version 4 (aggregate or cumulative). Prepared for August 14-15, 2007 OPP FIFRA Scientific Advisory Panel.

Glen G., Isaacs K., and Smith L. in preparation, Sobol's Method of Sensitivity Analysis Applied to Stochastic Models.

Glen G. 2010. SHEDS-Residential Version 4 SAS code.

Glen G., Smith, L., Isaacs, K., McCurdy, T., and Langstaff, J. 2008. A New Method of Longitudinal Diary Assembly for Human Exposure Modeling. *Journal of Exposure Science and Environmental Epidemiology* 18:299-311.

Graham S.E. and McCurdy T. 2003. Developing Meaningful Cohorts for Human Exposure Models. *Journal of Exposure Analysis and Environmental Epidemiology* 14(1):23-43.

Hore P., Zartarian V., Xue J., Özkaynak H., Wang S.W., Yang Y.C., Chu P.L., Sheldon L., Robson M., Needham L., Barr D., Freeman N., Georgopoulos P., and Lioy P.J. 2006. Children's Residential Exposure to Chlorpyrifos: Application of CPPAES Field Measurements of Chlorpyrifos and TCPy within MENTOR/SHEDS-Pesticides Model. *Science of the Total Environment* 366:525-537.

Isaacs K., Errickson A., Forbes S., Glen G., McCurdy L., McCurdy T., Smith L., Tulve N., and Vallero D. 2007. Statistical Properties of Longitudinal Time-Activity Data for Use in Human Exposure Models. Poster presented at the International Society of Exposure Analysis meeting, October 2007, Durham, NC.

Isaacs K., Glen G., McCurdy T., and Smith L. 2008. Modeling Energy Expenditure and Oxygen Consumption in Human Exposure Models: Accounting for Fatigue and EPOC. *Journal of Exposure Science and Environmental Epidemiology* 18:289-298

Isaacs K., McCurdy T., Errickson A., Forbes S., Glen G., Graham S., McCurdy L., Nysewander M., Smith L., Tulve N., and Vallero D. 2009. Statistical Properties of Longitudinal Time-Activity Data for Human Exposure Modeling. Poster presented at the American Time Use Research Conference, June 2009, College Park, MD.

Isaacs K., Stallings C., Zartarian V.G., and Glen G. 2010a. Stochastic Human Exposure and Dose Simulation (SHEDS) Model for Multimedia, Multipathway Chemicals: Version 4 Residential Module User Guide. Prepared for the July 20-22, 2010 EPA FIFRA SAP, Crystal City, VA.

Isaacs K., Xue J., Stallings C., and Zartarian V.G. 2010b. Stochastic Human Exposure and Dose Simulation (SHEDS) Model for Multimedia, Multipathway Chemicals: Version 1 SHEDS-Dietary Module User Guide. Prepared for the July 20-22, 2010 EPA FIFRA SAP, Crystal City, VA.

Isakov, V., Touma, J., Burke, J., Lobdell, D., Palma, T., Rosenbaum, A., and Ozkaynak, H. 2009. Combining Regional- and Local-Scale Air Quality Models with Exposure Models for Use in Environmental Health Studies. *Journal of Air & Waste Management Association* 59:461–472.

Johnson D., Thompson R., and Butterfield B. 1999. Outdoor Residential Pesticide Use and Usage Survey and National Gardening Association Survey. Unpublished study prepared by Doane Marketing Research, Inc. EPA MRID 46883825/44972202.

Keenan, J. 2007. Potential Exposures of Children and Adults to Cypermethrin and other Pyrethroid Insecticides Following Treatment and Control of Indoor Pests. (Doctoral Dissertation, University of California, Riverside, June 2007).

Kissel J.C., Shirai J.H., Richter K.Y., and Fenske R.A. 1998. Empirical Investigation of Hand-to-Mouth Transfer of Soil. *Bulletin of Environmental Contamination and Toxicology* 60: 379-386.

Leckie J.O., Naylor K.A., Canales R.A., Ferguson A.C., Cabrera N.L., Hurtado A.L., Lee K., Lin Y., Ramirez A., and Vieria J.D. 2000. Quantifying Children's Microlevel Activity Data from Existing Videotapes. Contract Report Submitted to US EPA, ORD, NERL, Reference Number U2F112OT-RT.

McCurdy T., Glen G., Smith L., and Lakkadi Y. 2000. The National Exposure Research Laboratory's Consolidated Human Activity Database. *Journal of Exposure Analysis and Environmental Epidemiology* 10:566-578.

National Gardening Association. 2009. The Impact of Home and Community Gardening in America; http://www.gardenresearch.com/home?q=show&id=3126

NHANES III. 1988-1994. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, (1988-1994) http://www.cdc.gov/nchs/nhanes/nh3data.htm

Nishioka M., Ivanovic W., Morara M., and Bortnick S. 2003. Characterizing Pesticide Residue Transfer Efficiencies: Report on Tracer Transfer Efficiencies to Hands Using Video-fluorescence Imaging. Report to U.S. EPA under Task Order 0017, Contract Number 68-D-99-011.

Özkaynak H., Xue J., Zartarian V., Glen G., Smith L. In Review. Modeled Estimates Of Soil And Dust Ingestion Rates For Children. *Risk Analysis*.

Shoaf M.B., Shirai J.H., Kedan G., Schaum J., and Kissel J.C. 2005. Child Dermal Sediment Loads Following Play in a Tide Flat. *Journal of Exposure Analysis and Environmental Epidemiology* 15:407–412.

Stahl W.R. 1967. Scaling of respiratory variables in mammals. *Journal of Applied Physiology* 22: 453-460.

Smith L. and Glen G. In preparation. Methods for correlating inputs to stochastic models.

Tsang A.M. and Klepeis N.E. 1996. A Detailed Analysis of the National Human Activity Pattern Survey (NHAPS) Respondents. Technical Report, EPA/600/R-96/148, US EPA, Washington D. C.

U.S. Census Bureau. 2002a. Census 2000 Summary File 3 – United States.

U.S. Census Bureau. 2002b. Census 2000 Summary File 3 Technical Documentation.

U.S. Dept. of Housing and Urban Development and US Census Bureau. 2007. American Housing Survey. http://www.census.gov/hhes/www/housing/ahs/ahs.html

U.S. EPA. 1993. Wildlife Exposure Factors Handbook. U. S. Environmental Protection Agency, Washington D.C., EPA/600/R-93/187.

U.S. EPA. 2006a. Organophosphate Cumulative Risk Assessment (2006 Update). DP Barcode: DP 331247. Office of Pesticide Programs. July 31, 2006.

U.S. EPA. 2006b. Permethrin: Third Revision of the Occupational and Residential Exposure Assessment for the Reregistration Eligibility Decision Document. Office of Pesticide Programs. April 4, 2006.

U.S. EPA. 2008. Child-Specific Exposure Factors Handbook. http://cfpub.epa.gov/ncea/CFM/recordisplay.cfm?deid=199243.

U.S. EPA. 2009. Standard Operating Procedures for Residential Pesticide Exposure Assessment, Draft Technical Guidelines. Office of Pesticide Programs. September, 2009.

U.S. EPA. 2010. Level I Quality Assurance Project Plan (QAPP) for the Stochastic Human Exposure and Dose Simulation Model for Multimedia, Multipathway Chemicals (SHEDS-Multimedia), National Exposure Research Laboratory, Human Exposure & Atmospheric Sciences Division, Prepared by Alion Science and Technology under Work Assignment #3-12 Contract No. EP-D-05-065.

Wong E.Y., Shirai J.H., Garlock T.J., and Kissel J.C. 2000. Adult proxy responses to a survey of children's dermal soil contact activities. *Journal of Exposure Analysis and Environmental Epidemiology* 10:509-17.

Xue J., McCurdy T., Spengler J., and Özkaynak H. 2004. Intra- and Inter-Individual Activity Considerations in Human Exposure Modeling. *Journal of Exposure Analysis and Environmental Epidemiology* 14: 221-233.

Xue J., Zartarian V., Özkaynak H., Dang W., Glen G., Smith L., and Stallings C. 2006 A Probabilistic Arsenic Exposure Assessment for Children Who Contact Chromated Copper Arsenate (CCA)-Treated Playsets and Decks: Part 2: Sensitivity and Uncertainty Analyses. *Risk Analysis* 26: 533-541.

Xue J., Zartarian V., Moya J., Freeman N., Beamer P., Black K., Tulve N., and Shalat S. 2007. A Meta-Analysis of Children's Hand-to-Mouth Frequency Data for Estimating Nondietary Ingestion Exposure. *Risk Analysis* 27: 411-20.

Xue J., Zartarian V., Tulve N., Moya J., Freeman N., AuYeung W., and Beamer P. 2010a. A Meta-Analysis of Children's Object-to-Mouth Frequency Data for Estimating Non-Dietary Ingestion Exposure. In Press. *Journal of Exposure Science and Environmental Epidemiology*.

Xue J., Zartarian V.G., and Nako S. 2010b. The Stochastic Human Exposure and Dose Simulation (SHEDS)-Dietary Model Technical Manual. Prepared for the July 20-22, 2010 EPA FIFRA SAP, Crystal City, VA.

Xue J., Zartarian V., Wang S-W., Liu S.V., and Georgopoulos P. 2010c. Probabilistic Modeling of Dietary Arsenic Exposure and Dose and Evaluation with 2003—2004 NHANES Data. *Environmental Health Perspectives* 118(3): doi:10.1289/ehp.0901205.

Young B., Driver J., Zartarian V., Xue J., Smith L., Glen G., Johnston J., Delmaar C., Tulve N., and Evans J. 2008. Dermal, Inhalation, and Incidental Exposure Results from the Models: How Did They Handle the Data? Paper presented at the joint meeting of International Society of Exposure Analysis and International Society of Environmental Epidemiology, October, 2008, Pasadena, CA.

Young B.M., Tulve N.S., Barnekow D., Delmaar C., Driver J., Egeghy P.P., Evans J., Glen G., Johnston J., Lunchick C., Miller D.J., Ross J.H., Smith L.A., Xue J., and Zartarian V.G. In preparation. Exposure Route Comparison of Four Probabilistic Aggregate Residential Exposure Models (SHEDS, CARES, Calendex, ConsExpo).

Zartarian V.G., Özkaynak H., Burke J.M., Zufall M.J., Rigas M.L., and Furtaw Jr. E.J. 2000. A Modeling Framework For Estimating Children's Residential Exposure and Dose to Chlorpyrifos Via Dermal Residue Contact and Non-Dietary Ingestion. *Environmental Health Perspectives* 108(6): 505-514.

Zartarian V.G., Xue J., Özkaynak H.A., Dang W., Glen G., Smith L., and Stallings C. 2005. A Probabilistic Exposure Assessment for Children Who Contact CCA-treated Playsets and Decks Using the Stochastic Human Exposure and Dose Simulation Model for the Wood Preservative Scenario (SHEDS-WOOD), Final Report. U.S. Environmental Protection Agency, Washington, DC, EPA/600/X-05/009.

Zartarian V., Xue J., Özkaynak H., Dang W., Glen G., Smith L., and Stallings C. 2006. A Probabilistic Arsenic Exposure Assessment for Children Who Contact CCA-Treated Playsets and Decks, Part 1: Model Methodology, Variability Results, and Model Evaluation. *Risk Analysis* 26: 515-531.

Zartarian V., Bahadori T., and McKone T. 2005. Feature Article: Adoption of an Official ISEA glossary. *Journal of Exposure Analysis and Environmental Epidemiology* 15: 1-5.

Zartarian V.G., Glen G., Smith L., and Xue J. 2008. SHEDS-Residential Model Version 3 Technical Manual, EPA 600/R-08/118, September 2008, http://www.epa.gov/heasd/products/sheds_multimedia/sheds_mm.html.

APPENDIX A Terminology Used in SHEDS-Residential

This appendix contains a list of terms used in this Technical Manual. The definitions here are not intended to be comprehensive or necessarily applicable outside the context of the SHEDS model; specific meanings in the context of the SHEDS model are presented for clarification to readers of this Technical Manual. Where possible, the SHEDS definitions are consistent with the International Society of Exposure Analysis glossary (Zartarian et al., 2005).

Term	Definition
Absorption	Absorption is the process resulting in *dose* for the dermal pathway. It results in the mass of chemical agent crossing the SHEDS exposure surface (which is the outer boundary of the skin for the dermal pathway) during a given CHAD diary event. The term is not used for the non-dietary ingestion and inhalation pathways; for those routes, the term intake is used.
Agent	A chemical of interest that contacts the human target and has potential for becoming a dose. SHEDS-Residential version 4 allows estimation of exposure and dose for multiple chemical agents.
Aggregate exposure	For a simulated individual, exposure from a single chemical summed over time and over contact media (and application scenarios, if applicable).
Air	Indoor air is a SHEDS contact medium in which the chemical agent may exist. Air concentrations are measured in units of ug/m3. Outdoors, air exposures can occur during the handling (application) of lawn and garden scenarios, but not thereafter.
Application	The release chemical agent into the residential environment. SHEDS allows multiple application scenarios in the same model run, each of which can be used (applied) multiple times by the same person.
Application scenario	Any chemical application type that leads to exposure (ex. crack-crevice, fogger, pet applications.) SHEDS runs may include multiple scenarios. Also called simply a "scenario."
Bathing	In SHEDS this is used as a generic term that includes baths, showers, and swimming. The effects of bathing are implemented at the end of the relevant diary events, reducing both body loading and hand loading. Bathing events can occur in two ways, either directly indicated on the activity diary, or added as supplemental information. See the discussion on the DiaryDetailing module for further details.

Binomial variable A binomial variable has only two possible outcomes. In SHEDS, these variables take the form of a "Yes/No" result. Version 4 requires the user to enter the Yes probability as the first parameter (v1), and the No probability as v2.

Body The Body is the entire skin surface except for the hands. The total body surface area for each individual is considered to be constant over time. At any given time, only a fraction of the body surface (re-sampled daily) is considered unclothed, which affects transfer rates.

CHAD The EPA's Consolidated Human Activity Database, described in McCurdy et al. (2000), and available online at http://www.epa.gov/chadnet1 . The original CHAD contained about 22,000 one-day human activity diaries, but has recently been expanded to around 34,000 diaries. CHAD is currently used in several EPA exposure models.

Cohort A group of individuals in SHEDS who share the same gender and belong to the same age category. For example, all 3-5 year old males may be called a cohort. The main relevance of cohorts is that the activity diaries are drawn from individuals in the activity database belonging to the same cohort as the individual being simulated. In some cases, input distributions may be designated by cohort. SHEDS output lists each person's specific age and gender, and so are not grouped into cohorts. The user may choose to define their own cohorts for summarizing the model output.

Contact event A CHAD diary event that involves new contact with a (potentially) contaminated contact medium.

Contact medium The contact media in SHEDS are the repositories for the chemical agent(s) before coming into contact with the simulated persons. There are nine contact media in SHEDS version 4: treated hard floors, treated carpets, untreated hard floors, untreated carpets, indoor air in treated areas, indoor air in untreated areas, lawn, vegetable garden, and pets.

Co-occurrence Relationships in usage patterns within and between scenarios. In SHEDS, within-scenario co-occurrence is exclusionary, meaning that multiple treatments cannot occur too near each other in time. Between-scenario co-occurrence may result in either enhanced or diminished probability for multiple scenarios being used together; for example, both the lawn and garden may be treated on or around the same date.

Concentration When plotted, a concentration time series has time on the x-axis and chemical

Time Series concentration on the y-axis. On SAS data sets, a time series is a single variable with one record for each point in time. Daily and hourly time series are spaced evenly in time, but SHEDS also has event-level time series which have varying-duration events.

Correlation The SHEDS user may request correlation between pairs of continuous input variables. This is done by selecting the pair and specifying the Spearman (or rank) correlation between them. The marginal distributions of each variable will remain exactly as specified elsewhere. For example, one might be triangular and the other a gamma. The method allows the two correlated variables in the pair to be sampled at different frequencies. The user may specify as many variable pairs as desired. Some combinations of pairwise correlations are contradictory and cannot be achieved. SHEDS determines when this occurs and informs the user.

Daily Dose This may refer to any of the DD_ variables in the SHEDS output, which all refer to daily doses measured in units of (mg/kg/day). For example, DD_Inhal is the daily inhalation dose. Depending on which data set this variable is found, it could represent the amount on a specific day, or the overall average for the person, or possibly the maximum of the daily values for that person.

Decay/dispersion This is one of the built-in methods for determining the change in chemical concentrations in SHEDS. The treated area undergoes first-order exponential decay in concentration, while for indoor scenarios nearby untreated areas may acquire chemical via dispersion from the treated area.

Dermal SHEDS separates the skin into two regions: hands and body. Although the transfer rates and loadings may differ, many SHEDS inputs are common to the two regions.

Diary event Diary events are the time basis on which all exposure and dose variables are calculated in SHEDS. Diary events have variable duration, from 1 to 60 minutes, and never cross the top of a clock hour. The number of diary events in the simulation period generally differs between individuals, because the CHAD diaries differ.

Distribution The expected set of outcomes for any randomly-sampled variable. Continuous variables may be sampled from beta, gamma, exponential, lognormal, normal, point, triangle, uniform, or Weibull statistical distributions. Probability vectors and binomial distributions are used for discrete outcomes. Empirical data may also be supplied, from which one point is selected at random each time the variable is re-evaluated. Distributions may be modified by correlating variables or by truncation of the extreme values. Any distribution may be replaced by a point value, which removes the random variation from that input.

Dose Dose is the mass of chemical agent crossing the exposure surfaces. At the event level, dose is measured in micrograms. Daily summaries are calculated, both in (ug/day) and in (mg/kg/day). SHEDS also calculates blood dose and urine dose at the daily level. These variables are reported only in units of (mg/kg/day) and refer to the amount of metabolite.

Exposure The amount of chemical agent present at the exposure surfaces (external skin surface, oral-nasal boundary). There are two measures of dermal exposure: new exposure only counts that which is picked up from the environment during the current diary event; and "running exposure", the total mass of chemical currently on the skin, resulting from all prior additions and subtractions. For air, exposure is the chemical concentration (in ug/m^3) being inhaled. For non-dietary ingestion, exposure is the amount of chemical placed in the mouth, regardless of whether it is removed (and becomes dose) or not.

Exposure surface The surface which must be crossed for "exposure" to become "dose". For the dermal pathways this is the outer surface of the skin. For non-dietary ingestion pathways it is the mouth. For inhalation exposure it is the oral-nasal boundary.

Exposure Time Series A set of values representing the exposure of one individual at a series of points in time, usually plotted with time on the x-axis and exposure on the y-axis.

Hands The *hands* refers to the skin area below the wrists on both hands, abbreviated "H" in certain variable names.

Hand washing Hand washing in SHEDS is randomly added at the end of diary events (except for sleeping events), using a probability that is specific to each individual. This probability is based on a user-specified input distribution for the mean number of hand washes per day. The result is that the particular times and the frequency of hand-washing events vary from day to day, even for the same individual. Hand washing reduces the hand loading, but does not affect the body or other loadings.

Individual One simulated person, for whom exposure and dose time series can be estimated.

Input variable An *input variable* is a quantity used in the determination of exposure or dose that is randomly sampled according to user-specified rules. It is stochastically sampled from one or more distributions. See also the *sampling frequency* and *job settings*.

Iterations	The total number of persons simulated in a model run. This is the product of the population size and the number of repetitions needed for uncertainty (or sensitivity) analysis. For a variability run, it is just the population size.
Micro	A short form for *microenvironment*, which in SHEDS indicates the individual's immediate surroundings for the duration of the current CHAD diary event. SHEDS uses six micros: in_home, out_home, in_other, out_other, in_vehicle, and sleep. The micro is determined from the CHAD location code (and the activity code in the case of sleep). Each activity diary event takes place in one micro, although for contact purposes there may be multiple contact media present.
New exposure	The amount of chemical agent transferred from an external contact medium onto the skin of the target person, on a given diary event. The term is used in distinction to *running exposure*. For the non-dermal pathways there is only one type of exposure, so the adjective *new* is not generally used.
Non-Chemical Specific Exposure Factors	Exposure factors only requiring a single input distribution (i.e.) the distribution does not vary by chemical
Non-dietary ingestion	Ingestion of chemical through hand or object mouthing.
Exposure pathways	The course a chemical agent takes from the source (or contact medium) to the target.
Population size	The number of individuals used to generate a variability distribution. This is distinct from the number of actual persons living in some area. SHEDS can handle any population size, subject to computer resource limitations.
Post-application	Occurring after (but not during) a residential chemical application or scenario usage event.
Probability vector	This is one of the alternatives for specifying the form of a stochastic input. It is used for variables that have more than two discrete outcomes, each of which is assigned a probability. The probabilities for the full set of outcomes should total to one. Probability vectors with just two possible outcomes are called *binomial distributions.*

Profile A simulated person having specific demographic and home characteristics

Running exposure *Running exposure* is the current mass of chemical on either the hands or the rest of the body. This includes both new exposure and the amount carried over from prior events. Running exposure is in micrograms (μg) of chemical, whereas *loading* is in (ug/cm^2).

Sampling frequency The *sampling frequency* determines the rate at which new samples are drawn for a given input variable. The most common sampling frequency is once per person. This means that each individual is given one value drawn at random, which remains fixed for that person. Other sampling frequencies include monthly, daily, or hourly. Such variables change values over time for a single person, as well as differing between persons.

Scenario Any chemical application type that leads to exposure (ex. crack-crevice, fogger, pet applications.) SHEDS runs may include multiple scenarios. Identical to "application scenario" They are used interchangeably.

Simulation period The *simulation period* is the time period for which exposures are estimated. This may range from one day to a year or more, measured in whole days. Each individual is modeled for the entire simulation period, which is the same for everyone. The exception to this is when the user specifies the *timeseries* method for concentrations and the input file contains more than one set of concentration measurements, with unequal measurement periods. In that case, each person is modeled only for the period with non-missing values in the set of measurements assigned to them.

Stochastic *Stochastic* means randomly determined or probabilistic. In SHEDS the sequence of persons and the values for many of the input variables are stochastic, so the individuals are independent and there is no significance to their order.

Target In SHEDS, the target is a simulated person that may receive an exposure.

Treated area An area in a simulated individual's residence which has been (or will be) treated at some point during the simulation. The definition does not change over time for one person, but may vary from person to person. Each indoor scenario may have a different treated area in version 4.

Uncertainty Refers to here is the lack of knowledge about the true value of a quantity, e.g. which of several alternative probability density functions should represent a quantity of interest. It can also represent a lack of knowledge about which of several alternative model representations best describes a biological, chemical, or physical mechanism of interest.

Untreated area The *untreated area* is the part of a simulated individual's residence which is not directly treated, but which may become contaminated by diffusion/dispersion from a treated area.

Variability Refers to the differences across the target population. The phrase "variability distribution" may be applied to both inputs and outputs; the reference will be clear from the context. The phrase "probability distribution" is occasionally used in the usual statistical sense to refer to general probability density functions. The phrase "probability vector" is reserved for discrete multinomial distributions.

APPENDIX B Selected SAS Variable Nomenclature

SAS variable	Description
age	age of simulated person in full years
area_body	skin surface area (except hands) in cm2
area_dermal	total skin surface area in cm2
area_garden	size of vegetable garden in m2
area_hands	surface area of both hands in cm2
area_house	floor area of house in m2
area_lawn	size of lawn in m2
area_pet	surface area of pet in m2
background	background chemical concentration on outdoor surfaces (ug/cm2)
blood_chemical	amount of chemical entering the blood per day (ug/day)
blood_metabolite	amount of new metabolite in blood (ug/day)
bmr	basal metabolite rate in megajoules per day
bva	basal alveolar ventilation rate in m3/hr
chadid	9-character CHAD ID code for 24-hour diary day
chemfrac	fraction of environmental chemical transferred to person
cohort	age-gender grouping for diary selection
day	day of the month (range 1-31)
day_num	sequential day number of the simulation period (first day=1)
DD_blood_chemical	daily blood dose of chemical (amount entering blood in mg/kg/day)
DD_blood_metabolite	daily blood dose of metabolite (chemical dose multiplied by mass ratio)
DD_dermal	daily dermal dose (amount absorbed into *stratum corneum* in mg/kg/day)
DD_ingest	daily ingested dose (amount of non-dietary chemical ingested in mg/kg/day)
DD_inhal	daily inhaled dose (amount of chemical inhaled in mg/kg/day)
DD_total	daily dose (total amount of chemical entering body in mg/kg/day)
DD_urine_metabolite	daily urine dose of metabolite (amount entering urine in mg/kg/day)
DE_dermal	daily dermal exposure (amount of chemical deposited on skin in mg/kg/day)
diaryA	one-day autocorrelation in key variable (new diary assembly method only)
diaryD	diversity statistic (ratio of between-person variance to total variance)
diaryT	personal mean ranking for key diary variable
dose_body	micrograms of chemical entering body via dermal pathway (excepting hands)
dose_dermal	micrograms of chemical entering body via dermal pathway (hands plus body)
dose_hands	micrograms of chemical dermally absorbed from hands
dose_HTM	micrograms of chemical transferred from hands to mouth
dose_ingest	micrograms of chemical from non-dietary ingestion (HTM plus OTM)
dose_inhal	micrograms of chemical inhaled into lungs
dose_OTM	micrograms of chemical transferred from objects to mouth
dose_total	micrograms of chemical entering body via all routes and pathways
dur_days	duration of event (days)
duration	duration of event (minutes)
dust_loading _hard	indoor dust loading on hard floors
dust_loading_soft	indoor dust loading on carpet (accessible dust only)
event_day	sequential event number (starting at one each day at midnight)

event_num	sequential event number within the entire simulation period
expo_body	new dermal exposure on the body (not hands) in ug
expo_dermal	new dermal exposure (hands plus body) in ug
expo_hands	new dermal exposure (hands only) in ug
expo_HTM	micrograms of chemical on hands entering mouth
expo_inhal	chemical concentration in inhaled air (ug/m3)
expo_OTM	micrograms of chemical on objects entering mouth
f_carpet	average fraction of indoor time spent on carpet
f_dermalblood	fraction of dermal dose that enters the blood
f_garden	average fraction of out-home time spent in garden
f_ingestblood	fraction of ingested dose that enters the blood
f_inhalblood	fraction of inhaled dose that enters the blood
f_lawn	average fraction of out-home time spent on lawn
f_pet	average fraction of at- home time spent with pet
f_treat_out	average fraction of out-home time in treated areas (lawn or garden)
f_urine	fraction of metabolite blood dose excreted in urine
f_uncloth	fraction of body unclothed
gender	gender of simulated person (M=male, F=female)
handler_body	dermal exposure to body (not hands) during scenario handling
handler_dermal	dermal exposure to hands and body during scenario handling
handler_hands	dermal exposure to hands during scenario handling
handler_inhal	inhalation dose during scenario handling
handwash_mean	mean number of hand washes per day
has_garden	vegetable garden status (1=residence has one, 0=does not)
has_lawn	lawn status (1=residence has one, 0=does not)
has_pet	dog or cat status (1=residence has one, 0=does not)
height	height of simulated person in cm
hour	hour of the day (range 1-24, 1=midnight to 1 a.m.)
hour_num	hour of the simulation (1=first hour of simulation period)
house	house number (used with TimeSeries input only)
InvMOE_dermal	reciprocal of dermal margin of exposure (MOE)
InvMOE_ingest	reciprocal of ingestion margin of exposure (MOE)
InvMOE_inhal	reciprocal of inhalation margin of exposure (MOE)
InvMOE_total	reciprocal of total margin of exposure (MOE)
mets_max	personal maximum allowed METS ratio
micro	SHEDS microenvironment for current diary event
MOE_dermal	dermal margin of exposure (MOE)
MOE_ingest	ingestion margin of exposure (MOE)
MOE_inhal	inhalation margin of exposure (MOE)
MOE_total	total margin of exposure (MOE)
month	month number (1=January, etc.)
person	number identifying each simulated person
postapp_body	postapplication dermal exposure on body (not hands) in ug
postapp_dermal	total postapplication dermal exposure in ug
postapp_hands	postapplication exposure on hands in ug
postapp_inhal	postapplication inhalation exposure in ug
reentry_in	reentry time in hours for indoor scenarios
reentry_out	reentry time in hours for outdoor scenarios
reentry_pet	reentry time in hours for pet scenarios

remove_B_bath	micrograms of chemical removed from body during bath/shower
remove_B_brushoff	micrograms of chemical removed from body by brush-off
remove_B_maxload	micrograms of chemical removed from body due to loading limit
remove_H_bath	micrograms of chemical removed from hands during bath/shower
remove_H_brushoff	micrograms of chemical removed from hands by brush-off
remove_H_maxload	micrograms of chemical removed from hands due to loading limit
remove_H_wash	micrograms of chemical removed from hands by hand washing
RunExp_body	running exposure on body (not including hands) in ug
RunExp_dermal	running dermal exposure in ug
RunExp_hands	running dermal exposure (hands only) in ug
seed_base	base random number seed for current person
seed_scenario	random number seed used for scenario variables
seed_simulation	random number seed used for simulation variables
start	start time for current diary event (minutes since midnight)
urine_metabolite	micrograms of metabolite excreted in urine
vent_inhal	breathing ventilation rate in (m3/min)
weight	body weight of simulated person in kg
year	calendar year

APPENDIX C SHEDS-Residential Code Modules and their Functions

While this Technical Manual does not attempt to be a programmer's guide, some knowledge of the organizing principles in the SHEDS code may be useful to the reader. The SHEDS code is written in SAS. All of the code (not including the GUI) is together is in a single file, but internally this code is divided into several modules.

Figure 2-1 illustrates the general structure of SHEDS-Residential version 4 and how the modules are linked. Conceptually, SHEDS is a two-stage stochastic model. Each pass through the inner loop produces exposure and dose time series for one simulated individual. Multiple passes produce results for a population. SHEDS focuses on exposure in and around one's home to chemicals that are commonly found in residential pesticides.

The SHEDS-Residential code consists of a main program and seven modules, and is described below. Each module is broken into several macros, which would be the equivalent of subroutines in certain other languages. The main program is quite short and is listed in full below. It contains three nested loops; the outermost one over chemical species, the next over either uncertainty or sensitivity iterations, and the innermost one over simulated individuals. Typically, each line in the main program calls one of the SHEDS macros. A few of these macros call another macro, but where possible just nesting has been avoided and never extends very far. With the exception of the utility macros, virtually all others are called from just one place in the code. This means that the division of the code into macros is largely a matter of conceptual divisions.

The first SHEDS module is the *Initialization module*, which reads the job settings, sets up the calendar for the simulation period, and determines the age, gender, and other properties of the simulated people. The macros in the initialization module are:

Macro	Purpose
%Start	Read job settings, prepare folders and work space.
%Calendar	Create the hourly and daily calendars for the simulation period.
%Chem_setup	Read the chemical-specific data for one chemical at a time.
%Population	Read the age-gender probability data and the cohorts to be modeled.
%OnePerson	Generate personal characteristics for one simulated person.

The last of the above steps requires random numbers that are provided by the Sampling module. The other steps refer only to the SHEDS input files. None of these macros calls other macros (except for Utility macros, which may be called from anywhere and are always excluded from comments like this).

The *Sampling module* reads the distributions and other rules for randomly sampling variables. It generates personal random number seeds, select points from uncertainty clouds, provides values for all the u_variables, and writes the transformation equations for these random variables to macro variables, so these equations may be evaluated where needed. In this respect, version 4 is organized very differently from version 3. The latter generated random variables whenever they were needed. Version 4 generates random variables only in the Sampling module, and in a reproducible manner. The remainder of the SHEDS code may use or transform these random samples, but cannot produce any additional randomness.

The macros in the Sampling module are:

Macro	Purpose
%Parm_Setup	Read distributions, sampling frequencies, conditions, and truncation points for the random variables.
%CorrelSetup	Write the equations for random variables that are correlated.
%StartLoop	Generate the 3 random number seeds for each simulated person.
%Parm_Eq	Write the transformation equations for all the u_variables.
%U_variables	Generate U(0,1) random samples for each random variable.
%U_Variables_Sobol	Generate pairs of U(0,1) random samples for each random variable.

The *Diary module* classifies the one-day activity diaries taken from the CHAD database and selects a set to represent each simulated person. Certain details such as hand-washing and bathing events may be added to the basic diaries. Finally, a METS level is assigned to each diary event, based on the CHAD activity. This is used later to determine the inhalation dose. The macros in the Diary module are:

Macro	Purpose
%DiaryPreProcess	Read CHAD diaries, accept or reject for use in SHEDS.
%DiaryGroups	Classify each diary into the appropriate cohort-pool.
%DiarySelection_Eight	Choose 8 diaries for each simulated person.
%DiarySelection _New	Use D & A method to construct longitudinal diary.
%DiaryDetailing	Add hand-washing, bathing, and METS.

The *Chemical module* is the largest one in SHEDS, in part because there are multiple options for specifying chemical usage and concentrations. This module produces time series of chemical concentrations in various contact media, weighted for potential contact duration and summed across scenarios. If the TimeSeries option is used, then the user supplies daily concentration time series directly. Otherwise, SHEDS first determines the usage patterns for the scenarios in the run. These usage patterns will be the same for all persons under the FixedDates option, but will be determined randomly for each person under the VariableDates option. With either choice, the resulting chemical concentrations may be determined using either the Decay/dispersion method or the Intervals method. The macros in the chemical module are:

Macro	Purpose
%TimeSeries	Read concentration time series data.
%Scenarios	Read list of scenarios and scenario-specific data.
%SoilProperties	Converts concentrations in soil to surface loadings.
%Dates_variable	Read rules for randomly assigning usage dates.

%UsageProbabilities	Determine usage probability for each day of the year.
%DetailOneUsage	Select hour and amount of chemical used.
%Dates_Fixed	Read list of usage dates common to all persons.
%Dates_None	Prepare dummy lists when no usage is specified.
%UsageHours	Adjust usage hour (if needed) and add handler information.
%Concentrations	Calculate chemical concentrations for each scenario and medium.
%ConcSummary	Adjust concentrations for contact duration and sum over scenarios.

The *Exposure module* brings together the outputs from the chemical and diary modules. The diary indicates where the person is located at each point in time. The concentrations reflect the amount of chemical present in the environment, at each location and point in time. The exposure is governed by transfer equations that involve randomly sampled transfer rates (from the stochastic module), durations (from the diaries), and concentrations (from the chemical module). The SHEDS exposure module determines the sequence of exposures, transfers, and doses, and then creates suitable summaries such as daily totals. The macros in the Exposure module are:

Macro	Purpose
%Exposure	Calculate time series of exposure by pathway.
%PersonSummary	Create daily summary variables for the current person.
%OuterLoopSummary	Summarize exposure and dose for a population.
%FinalSummary	Prepare text file for later use in a PBPK model.

The *Analysis module* consists of macros used in uncertainty or sensitivity runs. These are:

Macro	Purpose
%UncSummary	Draw two uncertainty-variability plots.
%SensPct	Create table of percentile scaling rankings.
%SensCorr	Correlate inputs with outputs.
%SensSobol	Evaluate Sobol sensitivity indices for groups of inputs.

The *Utility module* consists of the following macros:

Macro	Purpose
%ramp	Selects a time fraction for contact duration
%len	Counts the number of items in a list
%comma	Changes list delimiters from spaces to commas
%commaquote	Similar to %comma, but also puts each item in quotes

%removedups	Deletes repeated items from a list (if consecutive)
%locate	Finds the position of a selected item in a list
%addprefix	Adds a given prefix to every item in a list
%addsuffix	Adds a given suffix to every item in a list
%findprefix	Creates a sub-list of items containing a given prefix
%dropprefix	Like %findprefix, but it removes the prefix as well

The main SHEDS macro is shown below:

```
%macro Multimedia4(maindir,runname);
   %Start;
   %if &kill=1 %then %goto endmain;
   %Calendars;
   %if &DiaryPrep=1 %then %do; %DiaryPreProcess; %end;
   %do ChemNum=1 %to &NumChemicals;
      %Chem_setup;
      %if &UsageMethod=none %then %do; %TimeSeries; %end;
      %else %do; %Scenarios; %end;
      %Parm_setup;
      %do OuterLoop=1 %to &LastOuterLoop;
         %StartLoop;
         %Parm_eq;
         %Population;
         %DiaryGroups;
         %SoilProperties;
         %do InnerLoop=1 %to &LastInnerLoop;
            %U_variables;
            %OnePerson;
            %Dates_&UsageMethod;
            %UsageHours;
            %Concentrations;
            %ConcSummary;
            %DiarySelection_&DiaryMethod;
            %DiaryDetailing;
            %Exposure;
            %PersonSummary;
         %end;
         %OuterLoopSummary;
      %end;
   %end;
   %FinalSummary;
   %endmain:
%mend Multimedia4;
```

The main SHEDS program consists of three nested loops with a series of calls to other macros. Note that SAS code is not case-sensitive (except for string comparisons), so the capitalization is for the reader's convenience.

%Start reads the job settings and certain input data. If there are serious problems (for example, if the run directory cannot be found) then the variable &kill is set to one and SHEDS stops execution when %Start is finished. The %Calendars macro sets up the calendars for this run, using built-in SAS date

functions. SHEDS does not generally use year-sensitive inputs, so changing the year (without changing anything else) has only two effects: it may change the days of the week on which the simulation begins and ends, and it may add or remove a leap day (February 29).

The %DiaryPreProcess module is only run at special request. This is only required when the diary database is changed. For most purposes, the default diary database should be sufficient.

The outermost of the three nested %do loops is over chemicals. It would be possible to push this loop to the innermost position, but that would entail more changes from the version 3 code layout. In particular, the exposure and dose output files would contain data for multiple chemicals and would require variable names to identify which exposures belonged to each chemical. The current code layout avoids this problem by creating separate data sets on each pass through the chemical loop. The chemical name is attached as a suffix to these data sets, but is not present as part of the variable names within those data sets. A single chemical run uses the same code, only the chemical loop is executed just once. The %ChemSetup macro reads the input data specific to the chemical currently being modeled. Many SHEDS inputs (like concentrations and decay rates) are chemical-specific, so the distributions for these variables must be redefined on each pass through the chemical loop.

The next step depends on the user-selected chemical usage method, which may be Fixed Dates, Model-determined dates, or None. The third case requires that the user directly supply times series for the chemical concentrations. In the other two cases the %Scenarios macro is invoked to read the rules for determining the chemical usage dates.

%Parm_Setup reads all the distributions from which values are to be randomly selected. This includes the full list of random variables along with their sampling frequencies. If this is an uncertainty run, then the points randomly selected from the uncertainty clouds are chosen here. %Parm_Setup calls a specialized macro %CorrelSetup if the user has requested correlations between randomly sampled input variables.

The next %do loop is called the *Outer Loop*, which is a slight misnomer now that the main macro contains three loops. It refers to the outer stochastic sampling loop, which may be used for either uncertainty or sensitivity runs. (The loop over chemicals is not stochastic.) The outer loop is executed just once in a variability (single-stage) model run. The %StartLoop macro deletes data sets that need to be reinitialized on each pass through the outer loop, and it generates all the personal random seeds (three for each simulated person).

The %Parm_Eq macro generates the equations for producing random variables and writes them to SAS macro variables. This step must be repeated on each pass through the outer loop since the parameters for these input distributions may change. There are 9 SAS macro variables containing all the equations. Usually, one of these SAS macro variables is executed per data step, which means that the set of random variables on that macro variable are all evaluated as a group. For example, all the input variables relating to handlers are on the same SAS macro variable and will later on all be evaluated together.

%Population reads the age-gender data to determine the probabilities for each combination. This does not vary with each pass through the outer loop, so in principle it only needs to be determined once. However, it is a quick step and was therefore not moved from the position it inherited from the SHEDS version 3 code. %DiaryGroups assigns activity diaries to pools. Like %Population, this does not vary with each pass through the outer loop.

%SoilProperties is a new macro in version 4 that determines the conversion factor from chemical concentration in soil to an equivalent surface loading. This is based on the input distributions for various transfer variables along with the daily soil ingestion rate. SHEDS samples these distributions to

determine the appropriate conversion factor that makes the indirect soil ingestion rate (via hand loading and hand-to-mouth transfer) match the direct soil loading as specified by the user.

The innermost %do loop is over simulated individuals. Each pass through this loop is one person. In previous versions of SHEDS the person number was also the index number for this loop, but this was changed because SHEDS now permits the modeled of selected individuals from an earlier run. In that case, the loop indexing will differ from the person numbers.

The %U_variables macro produces all the u_variables (which are random samples from a U(0,1) distribution) that may be needed to model the current simulated person. These are stored until needed. For a Sobol sensitivity run, a variant of this macro (called %U_variables_Sobol) is invoked, which generates pairs of random numbers (often called the *sample* and *resample* values) instead of single values.

%OnePerson assigns the time-independent variables for the current individual. The most obvious ones are age and gender. Others include height, weight, skin surface area, basal ventilation rate, maximum ventilation rate (maxmets), and several other exposure-relevant variables, such as whether the household in question possesses a lawn, a garden, or a pet. SHEDS version 3 generated these variables for all the persons in the inner loop in a single data step, but version 4 generates just one person at a time. This is mainly due to random number control, as the Rand function in SAS permits setting of its seed value only once per data step. Since version 4 uses specific seeds for each person, only one person may be produced per data step. Later on, these records are appended onto a data set that covers all persons (which becomes the persons_demog output file).

The SAS macro variable &UsageMethod is set to Fixed, Variable, or None. Hence, one of the macros %Dates_Fixed, %Dates_Variable, or %Dates_None is invoked next. These macros assign scenario usage dates, if needed. The %UsageHours macro is skipped if the usage method is None; otherwise it produces an hourly data set called *apphours* that contains all the chemical usage information for the current person.

The %Concentrations macro is one of the main components of SHEDS. Along with the usage macros discussed immediately above and the %ConcSummary macro discussed below, these constitute the SHEDS concentration module. This module produces the effective chemical concentration time series for all potential contact media, taking into account the chemical usage patterns, all relevant application scenarios, and the contact probabilities. The %Concentrations macro produces one SAS data set for each scenario; each one contains hourly concentration time series for each contact medium. The %ConcSummary macro factors in the contact time fractions from the various calls to the ramp function and summarizes the results across application scenarios. The output consists of hourly time series for each of the following: hard floors, carpet, air, lawn, garden, pet, and sleep.

The SAS macro variable &DiaryMethod is set to either Eight or New. %DiarySelection_Eight chooses eight one-day activity diaries per person using the method from SHEDS version 3. %DiarySelection_New uses the new longitudinal diary assembly method described in section 2.5, using D and A statistics. The %DiaryDetailing macro adds hand-washing and bath events (if needed), and determines the Mets value for each diary event. The Mets value will be used in the exposure module to determine the inhaled dose. The diary module produces a time series covering the entire simulation period (often taken to be one year), broken into a series of “events” which last from 1 to 60 minutes each. A typical year-long diary may contain 15,000 – 20,000 events.

%Exposure is the heart of the SHEDS code. This macro combines the activity diary information with the chemical concentrations to obtain exposure, and subsequently dose. The exposure code in version 4 is simpler and easier to read than in version 3 for two reasons: first, the amount of macro language in the code has been reduced, and second, the contact duration logic has been moved to the

concentration module. Even so, there are a lot of computations of each diary event and the %Exposure macro remains one of the longer ones in SHEDS.

The %PersonSummary macro creates daily summaries for all the relevant exposure variables, and produces the blood and urine dose estimates for the current simulated person. These are appended to large files covering all persons simulated so far, before proceeding to the next pass through the inner loop.

%OuterLoopSummary calculates population averages. One record is added to the loops_mean_X output file for each pass through the outer stochastic sampling loop.

%FinalSummary is the last macro in SHEDS_Multimedia and it performs three tasks. First, it calls specialized analysis routines for sensitivity or uncertainty runs. Second, it creates the ASCII text files for use by a PBPK model, Third, it prints job timing information.

APPENDIX D Summary of the CHAD Database

The Consolidated Human Activity Database (CHAD) was originally created by compiling 12 studies into a single human activity pattern database resulting in 22,968 diary days (McCurdy et. al., 2000). Recently, several additional studies have been added, bringing the number of diary days up to 34,773.

As noted earlier in the manual (see Table 3-5), certain criteria were applied to CHAD to eliminate diaries that are unsuitable for human exposure modeling. For example, diaries with unspecified age or gender were dropped. The SHEDS installation package contains 33,748 diaries, but 587 of these have age=0 and cannot currently be used by the model, so effectively the database contains 33,171 useable diaries. Table D-1 lists the studies, their dates, age ranges, and the number of diaries available for SHEDS.

Table D-1. Summary of CHAD diaries

Study	Dates	Ages	CHAD Diary Count	SHEDS default diary count
Baltimore	1997-1998	72-93	391	380
California: Adults	1987-1988	18-94	1,579	1,541
California: Children	1989-1990	0-11	1,200	1,184
California: Youth	1987-1988	12-17	183	180
Cincinnati	1985	0-86	2,614	2,553
Denver	1982-1983	18-70	805	753
EPA longitudinal	1999-2008	35-67	1,427	1,419
Los Angeles: Elementary School	1989	10-12	51	47
Los Angeles: High School	1990	13-17	43	37
National: NHAPS Air	1992-1994	0-93	4,723	4,592
National: NHAPS Water	1992-1994	0-93	4,663	4,555
Ozone Averting Behavior	2002-2003	2-12	2,907	2,691
RTP Panel	2000-2001	55-85	1,000	873
PSID 1997: Children	1997	0-13	5,616	5,570
PSID 2002: Children	2002-2003	5-19	4,782	4.755
Seattle	1999-2002	6-91	1,693	1,644
Valdez, Alaska	1990-1991	11-71	397	308
Washington, DC	1982-1983	18-98	699	666
All Studies	1982-2008	0-98	34,773	33,748

SHEDS groups activity diaries into 10 age categories. Table D-2 provides the age-gender breakout of CHAD diaries available to SHEDS.

Table D-2. Diary Counts for SHEDS Age-Gender Cohorts in the default database.

Age / Gender	1-2	3-5	6-12	11-19	21-49	50-99
Female	858	1,355	4,250	1,809	4,467	4,111
Male	943	1,612	4,850	1,696	3,664	3,546

While the above numbers may seem large enough to avoid any sample size issues, note that SHEDS divides each of these cohorts into eight day-type pools by season and weekend-weekday. Since the studies that make up the CHAD database were not all year-round, sometimes certain pools are smaller than others. The smallest pool in the default diary database has 17 diaries, and there are 5 pools (out of 96) that have fewer than 50 diaries.

Table D-3 provides an example of a CHAD diary as it is provided to the SHEDS code. This is contained in the first column of the table. The two text columns are simply for descriptive purposes here; they are not a part of the file supplied to SHEDS. There is one record in this file for each event reported in the CHAD diary. As indicated in the table, not all diary event are shown in this example.

The variables in the first columns are separated by commas. In order, they are: CHAD ID code, event start time (in military time), event duration in minutes, CHAD activity code, and CHAD location code. The final comma in each event in the list is a CHAD artifact and of no consequence.

Table D-3. Example portion of CHAD Diary.

Contents of CHAD File	Activity	Location
LAE01606A,0000,60,14500,30120,	sleep	home, indoor
...		
LAE01606A,0700,11,14500,30120,	sleep	home, indoor
LAE01606A,0711, 4,14600,30120,	dressing	home, indoor
LAE01606A,0715, 5,14700,30120,	personal care	home, indoor
LAE01606A,0720,23,14400,30120,	eat	home, indoor
LAE01606A,0743,16,14700,30120,	personal care	home, indoor
...		
LAE01606A,0819, 5,17250,35000,	relax, wait	other outdoor
LAE01606A,0824, 6,14700,30120,	personal care	home, indoor
LAE01606A,0830,30,14400,31110,	eat	car
LAE01606A,0900,18,14400,31110,	eat	car
LAE01606A,0918,15,16210,33700,	religion	church
LAE01606A,0933,27,17180,33700,	active leisure	church
...		
LAE01606A,2300,60,14500,30120,	sleep	home, indoor

Table D-4 provides an exhaustive list of CHAD activity codes and their descriptions.

Table D-4. CHAD Activity Codes.

Code	Code Description
10000	Work and other income producing activities, general
10100	Work, General
10110	Work, general, for organizational activities
10111	Work for professional/union organizations
10112	Work for special interest identity organizations
10113	Work for political party and civic participation
10114	Work for volunteer/ helping organizations
10115	Work of/for religious groups
10116	Work for fraternal organizations
10117	Work for child / youth / family organizations
10118	Work for other organizations
10120	Work, income-related only
10130	Work, secondary (income-related)
10200	Unemployment
10300	Breaks
11000	General household activities
11100	Prepare food
11110	Prepare and clean-up food
11200	Indoor chores
11210	Clean-up food
11220	Clean house
11300	Outdoor chores
11310	Clean outdoors
11400	Care of clothes
11410	Wash clothes
11500	Build a fire
11600	Repair, general
11610	Repair of boat
11620	Paint home / room
11630	Repair / maintain car
11640	Home repairs
11650	Other repairs
11700	Care of plants
11800	Care for pets/animals
11900	Other household
12000	Child care, general
12100	Care of baby
12200	Care of child
12300	Help / teach
12400	Talk /read
12500	Play indoors
12600	Play outdoors
12700	Medical care-child
12800	Other child care

13000	Obtain goods and services, general
13100	Dry clean
13200	Shop / run errands
13210	Shop for food
13220	Shop for clothes or household goods
13230	Run errands
13300	Obtain personal care service
13400	Obtain medical service
13500	Obtain government / financial services
13600	Obtain car services
13700	Other repairs
13800	Other services
14000	Personal needs and care, general
14100	Shower, bathe, personal hygiene
14110	Shower, bathe
14120	Personal hygiene
14200	Medical care
14300	Help and care
14400	Eat
14500	Sleep or nap
14600	dress, groom
14700	Other personal needs
15000	General education and professional training
15100	Attend full-time school
15110	Attend day-care
15120	Attend K-12
15130	Attend college or trade school
15140	Attend adult education and special training
15200	Attend other classes
15300	Do homework
15400	Use library
15500	Other education
16000	General entertainment / social activities
16100	Attend sports events
16200	Participate in social, political, or religious activities
16210	Practice religion
16300	Watch movie
16400	Attend theater
16500	Visit museums
16600	Visit
16700	Attend a party
16800	Go to bar / lounge
16900	Other entertainment / social events
17000	Leisure, general
17100	Participate in sports and active leisure
17110	Participate in sports
17111	Hunting, fishing, hiking

17112	Golf
17113	Bowling / pool / ping pong / pinball
17114	Yoga
17120	Participate in outdoor leisure
17121	Play, unspecified
17122	Passive, sitting
17130	Exercise
17131	Walk, bike, or jog (not in transit)
17140	Create art, music, participate in hobbies
17141	Participate in hobbies
17142	Create domestic crafts
17143	Create art
17144	Perform music / drama / dance
17150	Play games
17160	Use of computers
17170	Participate in recess and physical education
17180	Other sports and active leisure
17200	Participate in passive leisure
17210	Watch
17211	Watch adult at work
17212	Watch someone provide childcare
17213	Watch personal care
17214	Watch education
17215	Watch organizational activities
17216	Watch recreation
17220	Listen to radio / listen to recorded music / watch T.V.
17221	Listen to radio
17222	listen to recorded music
17223	Watch TV
17230	Read, general
17231	Read books
17232	Read magazines / not ascertained
17233	Read newspaper
17240	Converse / write
17241	Converse
17242	Write for leisure / pleasure / paperwork
17250	Think and relax
17260	Other passive leisure
17300	Other leisure
18000	Travel, general
18100	Travel during work
18200	Travel to/from work
18300	Travel for child care
18400	Travel for goods and services
18500	Travel for personal care
18600	Travel for education
18700	Travel for organizational activity

18800	Travel for event / social activity
18900	Travel for leisure
18910	Travel for active leisure
18920	Travel for passive leisure
U	Unknown
X	Missing

Table 1-5 provides an exhaustive list of CHAD location codes and their descriptions.

Table 1-5. CHAD Location Codes.

Code	Code Description
30000	Residence, General
30010	Your Residence
30020	Other's Residence
30100	Residence, indoor
30120	Your residence, indoor
30121	Kitchen
30122	Living room / family room
30123	Dining room
30124	Bathroom
30125	Bedroom
30126	Study / Office
30127	Basement
30128	Utility room / Laundry room
30129	Other indoor
30130	Other's residence, indoor
30131	Other's Kitchen
30132	Other's living room / family room
30133	Other's Dining room
30134	Other's Bathroom
30135	Other's Bedroom
30136	Other's Study / Office
30137	Other's Basement
30138	Other's utility room / laundry room
30139	Other indoor
30200	Residence, Outdoor
30210	Your residence, Outdoor
30211	Your residence - Pool, spa
30219	Your residence - Other outdoor
30220	Other's residence, outdoor
30221	Other's residence - Pool, spa
30229	Other's residence - Other outdoor
30300	Garage
30310	Indoor garage
30320	Outdoor garage
30330	Your garage
30331	Your indoor garage
30332	Your outdoor garage
30340	Other's garage
30341	Other's indoor garage
30342	Other's outdoor garage
30400	Other, residence
31000	Travel, general
31100	Motorized travel

31110	Travel by car
31120	Travel by truck
31121	Travel by truck(pick-up van)
31122	Travel by Truck (other than pick-up or van)
31130	Travel by Motorcycle / moped / motorized scooter
31140	Travel by bus
31150	Travel by Train / Subway / rapid transit
31160	Travel by airplane
31170	Travel by boat
31171	Travel by motorized boat
31172	Travel by unmotorized boat
31200	non-motorized travel
31210	Travel by walk
31220	Travel by bicycle / skateboard / roller-skates
31230	Travel in a stroller or carried by an adult
31300	Waiting
31310	Wait for bus, train, ride (at stop)
31320	Wait for travel, indoors
31900	Other travel
31910	Travel by other vehicle
32000	Other, indoor general
32100	Office building / bank / post office
32200	Industrial plant / factory / warehouse
32300	Grocery store / convenience store
32400	Shopping mall / non-grocery store
32500	Bar / night club / bowling alley
32510	Bar / Night Club
32520	Bowling alley
32600	Repair shop
32610	Auto repair shop /gas station
32620	Other repair shop
32700	Indoor gym / sports or health club
32800	Childcare facility
32810	Childcare facility, house
32820	Childcare facility, commercial
32900	Public building / library / museum / theater
32910	Auditorium, sport's arena / concert hall
32920	Library / courtroom / museum / theater
33100	Laundromat
33200	Hospital / health care facility / doctor's office
33300	Beauty parlor / barber shop / hair dresser's
33400	At work : no specific location, moving among locations
33500	At School
33600	At Restaurant
33700	At Church
33800	At Hotel /Motel
33900	At Dry cleaners

34100	Parking garage
34200	Laboratory
34300	Other, indoor
35000	Other outdoor, general
35100	Sidewalk / street / neighborhood
35110	Within 10 yards of street
35200	Public garage / parking lot
35210	Public garage
35220	Parking lot
35300	Service station / gas station
35400	Construction site
35500	Amusement park
35600	School grounds / playgrounds
35610	School grounds
35620	playground
35700	Sports stadium and amphitheater
35800	Park /golf course
35810	Park
35820	Golf course
35900	Pool, river, lake
36100	Restaurant, picnic
36200	Farm
36300	Other outdoor
U	Uncertain
X	Missing

APPENDIX E Probability Density Functions

The explicit probability density functions (pdfs) utilized by SHEDS are listed below. Note that some of these may have alternate parametrizations, so the user must be careful when obtaining distributions from the literature. The expressions "Exp," "Sqrt," "Log," and "Γ" refer to the exponential, square root, natural logarithm, and gamma functions, respectively.

E.1 Beta

The beta distribution in SHEDS has a lower bound of zero, an upper bound of one, and two shape parameters v1 and v2. The restrictions are v1>0 and v2>0. When v1< v2 then the mean is below ½ and the distribution is positively skewed, whereas when v1 > v2 the mean is above ½ and the distribution is negatively skewed. For v1=v2 the mean is at ½ and the shape is symmetric. The PDF is

Equation E-1

$$p(x) = x^{v1-1} (1-x)^{v2-1} \Gamma(v1+v2) / (\Gamma(v1) \Gamma(v2)), \text{ for } 0<x<1$$

The beta is a useful form for variables known to be bounded, due to the wide variety of shapes that it can have. For v1>1 and v2>1, the PDF of the beta has a single peak, away from the bounds. When 0<v1<=1, the PDF is large near zero, and when 0<v2<=1, it is large near one. These properties allow the beta to have the so-called 'J' or 'U' shapes. The mean of a beta distribution is at $\mu = v1 / (v1+v2)$, and the standard deviation is $\sigma = \text{Sqrt}[v1\ v2 / (v1+v2+1)] / (v1+v2)$. If one wishes to construct a beta with a given mean μ and standard deviation σ, then choose $v1 = (\mu^2-\mu^3)/\sigma^2 -\mu$, and $v2 = v1 (1-\mu) /\mu$. This will only be possible if $0<\mu<1$ and $\sigma^2 < \mu (1-\mu)\leq 1/4$.

Some users may be familiar with four-parameter beta distributions in which the locations of the two bounds are variable as well. Suggestions have been made to implement these in SHEDS. This has not been adopted for two reasons: first, most of the bounded SHEDS variables have natural limits of zero and one (for example, any proper fraction such as the dermal loading reduction during hand washing), and second, SHEDS currently can accommodate only three parameters per distribution. Changing this to four parameters would require code and interface changes that have not yet been deemed worthwhile.

E.2 Exponential

The exponential in SHEDS has two parameters, the minimum (v1) and the mean (v2), with the restriction that v1 < v2. Some users may be more familiar with a single parameter exponential distribution, which has a minimum of zero and is characterized by a decay rate constant. The SHEDS exponential is similar, apart from a shift of v1 units to the right. The decay rate of the SHEDS exponential is given by 1/(v2-v1). The standard deviation of an exponential is (v2-v1). If the user wants an exponential with a half-life τ, then set $v2 = v1+ \tau / \text{Log}[2]$. The PDF of the SHEDS exponential is

Equation E-2

$$p(x) = \text{Exp}[-(x-v1)/(v2-v1)] / (v2-v1), \quad \text{for } x>v1$$

An exponential can also be obtained by requesting a gamma or a Weibull distribution with shape parameter set to one. However, in those cases the minimum is automatically at zero, as three-parameter forms of the gamma, Weibull and lognormal have not been implemented in SHEDS.

E.3 Gamma

The gamma distribution in SHEDS is bounded below by zero and has two parameters, the shape parameter v1 and the scale parameter v2. The restrictions are v1>0 and v2>0. The shape parameter v1 controls the appearance of the PDF. Shape parameters less than one lead to a monotonically decreasing form with the highest probability at zero. If v1=1, then the gamma is identical to an exponential that starts at zero and has a mean given by the gamma parameter v2. If v1>1, then the gamma somewhat resembles the lognormal, rising from zero to a peak probability, and then gradually declining with an overall positive skewness. The mean of the gamma is at $\mu = v1\ v2$, and the standard deviation is $\sigma = v2\ \mathrm{Sqrt}(v1)$. The PDF of the SHEDS gamma is

Equation E-3

$$p(x) = v2^{-v1}\ x^{v1-1}\ \mathrm{Exp}(-x/v2) / \Gamma(v1), \quad \text{for } x>0$$

The gamma, like several of the other distributions, is open-ended to the right. This means that regardless of the parameters used, the random sampling may occasionally produce extremely large numbers. The argument that a distribution is gamma-shaped is usually based on a relatively small number of observations, in which case very few will be far out in the tail. There may be practical reasons for an upper limit on the returned values. If so, then the user should truncate the distribution at such a limit. If the truncation point is far out in the tail, then truncation will rarely occur and will not appreciably alter the statistical properties of the distribution (such as its mean and variance). However, such truncation may prevent the generation of unreasonable values for the modeling variables.

E.4 Lognormal

The lognormal in SHEDS is bounded below by zero and has two parameters, the geometric mean GM (v1) and the geometric standard deviation GSD (v2). The restrictions are v1>0 and v2>1. Many variables in exposure science are approximately lognormally distributed, so its use is fairly common. If a variable *x* has a lognormal distribution, then log(x) has a normal distribution.

The geometric mean (GM) of a lognormal distribution is also its median. Log(GM) is the mean of the distribution of log(x). Log(GSD) is the standard deviation of log(x). Since standard deviations must be positive, then Log(GSD)>0, which implies GSD>1. The PDF of the SHEDS lognormal is

Equation E-4

$$p(x) = \mathrm{Exp}[(-1/2)\ (\mathrm{Log}[x/v1]\ /\mathrm{Log}[v2])^2\ ]\ /\ (x\ \mathrm{Sqrt}[2\ \pi\ ]\ \mathrm{Log}[v2]\ ), \quad \text{for } x>0$$

If GM and GSD are given, then the lognormal has the following arithmetic mean and standard deviation

Equation E-5

$$\mu = GM\ \mathrm{Exp}[(1/2)\ (\mathrm{Log}(GSD))^2\ ],$$

$$\sigma = GM\ \mathrm{Sqrt}[\mathrm{Exp}(\ [\mathrm{Log}(GSD)]^2)\ (\mathrm{Exp}(\ [\mathrm{Log}(GSD)]^2)-1)].$$

If the user knows the arithmetic mean μ and arithmetic standard deviation σ of the lognormal instead of the GM and GSD, then these can be converted as follows:

Equation E-6

$$GM\ = \mu\ /\ \mathrm{Sqrt}(1 + \sigma^2\ /\mu^2),$$

GSD = Exp(Sqrt(Log(1 + σ^2/μ^2))).

If instead, one has the mean μ_{log} and standard deviation σ_{log} of the logarithm log(x) of the underlying variable, then use

Equation E-7

GM = Exp(μ_{log}),

GSD = Exp(σ_{log}).

Like the gamma, the lognormal is open-ended to the right and it may be desirable to truncate the tail to prevent the possibility of unreasonably large numbers being returned.

E.5 Normal

This is the normal or Gaussian distribution commonly used in statistics. The normal has two parameters: the mean (v1) and the standard deviation (v2), with v2>0. Note that the normal is unbounded, so it is a good idea to provide lower and upper truncation points to prevent physically impossible values from being returned. The PDF of the normal is

Equation E-8

p(x) = Exp[-(x-v1)^2/(2 v2^2)] / (Sqrt[2 π] v2)

Truncation is especially important for the normal because the distribution may return negative values, and these are almost always meaningless in the context of the SHEDS modeling variables.

E.6 Point

A point value means that the same value is always returned. This is sometimes called a *fixed* or *constant* form. The point has one numeric argument (v1) which is the value that is to be returned. The mean is v1 and the standard deviation is zero. The sampling frequency does not matter for points. While points are technically discrete, here they are classified with the continuous distributions since they are applied to variables that are expected to reside on a continuous scale, but happen to be assigned no variability.

E.7 Triangular

The triangular distribution has a probability density function (PDF) that is shaped like a triangle. The three parameters locate the vertices, with v1=minimum, v2=peak, v3=maximum. The restrictions are v1<= v2 <= v3, with v1<v3. The mean value of this distribution is located at μ = (v1+v2+v3)/3, which coincides with the peak only when v2 is midway between v1 and v3. The standard deviation is σ = Sqrt[(v1^2+v2^2+v3^2-v1v2-v1v3-v2v3)/18]. It is possible for the peak to be located at either extreme, forming a right triangle. The PDF of the triangular is

Equation E-9

p(x) = 2 (x-v1) / [(v2-v1) (v3-v1)], for v1$\leq$ x $\leq$ v2

$= 2 (v3-x) / [(v3-v2) (v3-v1)], \quad \text{for } v2 \le x \le v3$

The SHEDS code only scans the first four characters of the distributional form, so this may be called either a *triangular* or *triangle* distribution.

E.8 Uniform

The uniform is characterized by two parameters, the minimum (v1) and the maximum (v2), with v1<v2. All values between v1 and v2 are equally likely to be returned. The mean is $\mu = (v1+v2)/2$ and the standard deviation is $\sigma = (v2-v1) / \text{Sqrt}(12)$. The PDF of the uniform is

Equation E-10

$$p(x) = 1 / (v2-v1), \quad \text{for } v1 < x < v2.$$

E.9 Weibull

The Weibull distribution in SHEDS is bounded below by zero and has two parameters, the shape parameter v1 and the scale parameter v2. The restrictions are v1>0 and v2>0. The Weibull has slightly different properties from a gamma, but there is a strong overall resemblance. When the shape parameter $v1 \le 1$, the Weibull is monotonically decreasing. For v1=1, it reduces to an exponential. For v1>1, it rises to a peak and then declines gradually in a long tail. The mean is $\mu = v2\ \Gamma(1+1/v1)$, and the standard deviation is $\sigma = v2\ \text{Sqrt}[\Gamma(1+2/v1) - (\Gamma(1+1/v1))^2]$. Here '$\Gamma$' is the mathematical gamma function, not the gamma distribution. The gamma function is a generalization of the factorial function to non-integer arguments; for integers, $\Gamma(1+n) = n!$. The PDF of the Weibull distribution is

Equation E-11

$$p(x) = v1\ v2^{-v1}\ x^{v1-1}\ \text{Exp}\ [-(x/v2)^{v1}], \quad \text{for } x>0.$$

E.10 Discrete Probability Density Functions

Binomial variables have only two possible outcomes, for example, the outcome of yes/no tests. The usual statistical notation would be (p_1, p_2), where p_1 and p_2 are between 0 and 1 and sum to 1. However, the SHEDS code implements binomial variables by having the user specify only v1, which is the probability of *yes*. The probability of no is implied as (1 - yes). The form is indicated as *binomial*.

Multinomial variables allow more than two possible outcomes. The usual statistical notation would be $(p_1, p_2, \ldots, p_n)$, where each p_i is between 0 and 1 and the p_i's sum to 1. In SHEDS, multinomial variables are called probability vectors. Unlike Bernoulli variables, the user must supply the entire set of probabilities. The form is indicated by *probability*. In this case the parameters v1, v2, and v3 are not used; instead the entire vector is presented as a comma-separated list of values under the *probvector* column on the input file. The SHEDS interface handles this task automatically.

APPENDIX F Description of SHEDS-Residential Input Files

F.1 List of Input Files

The following are the key input data files that the user works with to run the SHEDS-Residential version 4.0 model:

agegroups:	This contains the age and gender groups to be included in the model run.
areatreated:	This file provides time fractions spent in treated areas for the indoor application scenarios, which are used to determine the contact probabilities.
correlations:	This contains pairs of stochastic input variables and the desired Spearman correlation between them.
dates_fixed:	This lists the application scenario application types and the rules for having the model randomly determine the dates when these occur. Each modeled person is likely to have different usage dates.
dates_variable:	This lists the application scenario types and the specific dates on which these occur, with the same dates for all persons.
decay:	A list of application scenario-specific input variables and their distributions, which include decay rates and chemical concentrations immediately following an application.
distributions:	This contains one or more probability distributions for each variable listed in the variables file with userdistrib=1, and the conditions indicating when each of these distributions should be used.
intervals:	This replaces the decay file, if the user prefers to specify distributions for chemical concentrations in terms of the amount of time that has passed since application.
scenarios:	This contains the list of scenario application types to be included in the model run. This list is data-driven, and more scenarios may be added without altering the SHEDS code. However, if the list is expanded here, then the other input files that contain scenario-specific data would have to be expanded as well.
timeseries:	This file contains daily chemical concentration time series data for the relevant contact media. If this option is used, no scenarios or usage data are required.
variables:	This contains names, units, and descriptions of personal input variables, exposure factors, dose factors, and the re-sampling frequencies for these variables.

In addition, the following files are used internally by the model. These are applicable for a wide range of application scenarios and may be used without alteration in most cases:

bwsa:	This file contains distributions for body weight, height, and basal metabolic rate, as functions of age and gender.
cohorts:	This file is used to group the activity diaries into pools.
contactmedia:	This contains a list of the possible contact media in SHEDS.
diaryevents:	This is the large file containing all the diary events from all persons in the CHAD database.
diarylocs:	This file contains the mapping from CHAD location codes to the microenvironments used in SHEDS.
diaryQArules:	This file contains a list of logical rules for accepting a CHAD diary for use in SHEDS-Residential 4.
diaryquest:	This file contains the personal data from CHAD that are not changing over the course of the diary day, like age, gender, home state, diary date, day of week, season, temperature, and many other variables.
metsdists:	This file contains the activity-specific distributions for METS, which is the multiple of the basal metabolic rate needed to sustain the current level of activity.
metsmap:	This is a mapping of CHAD activity codes to METS distributions.
pop2000:	This file contains an age-gender summary of the 2000 U.S. Census, without any geographical detail.

Files containing concentration time series may also be input, if the user requests it. For formatting details, see the chapter on model inputs.

F.2 Detailed Description of User-Specified Input Files

Each input file is a SAS data set, which is organized like a spreadsheet. The columns are called *variables* and the rows are *observations* or *records*. Each variable has a format, either a character string or numeric. Character strings are indicated by "$n", where *n* is the maximum number of characters the variable may hold. Numeric variables are indicated below by "num".

The number of records on each file is flexible in SAS. Records may therefore be added or deleted without changing the code. However, additional records will have no effect unless some reference is made to them. Deleted records may cause missing values to be generated, if references to these are encountered in the code.

F.2.1 Variables file

The *variables* file contains background information on most of the stochastic input variables. It does not normally require changes from the user, and cannot be altered using the SHEDS interface. The *variables* file contains the SAS variables listed in Table F-1:

Table F-1. Column definitions for the variables file.

Column name	Format	Purpose
VarName	$24	name of input variable, as appearing in the code
Priority	num	order in which the records are to be processed
VarLabel	$80	description of the input variable
Units	$16	units for the input variable
SampFreq	$8	the resampling frequency
UserDistrib	num	flag for file on which data appears
VarType	$12	Indicates which data step generates random numbers
VarGroup	$8	the macro variable containing the equations for this input
SobolGroup	num	group number (used only with Sobol analysis)
SensPct	num	flag for use with percentile scaling sensitivity run

The units are for informative purposes only; the code makes no use of them. The priority is normally not important, but the order of evaluation could matter if the descriptions file makes use of conditions that refer to other input variables. The SampFreq variable is important since it determines the resampling frequency. This can be altered for any Simulation variable, but not any others. The last two columns are only used in specialized sensitivity analysis runs.

The variables file contains 70 records plus 3 per application scenario, which amounts to 97 records for the default version.

F.2.2 Distributions file

The *Distributions* file contains most of the distributions from which the input variables are to be randomly sampled. The exceptions are the application scenario-specific inputs, and population, census, or other demographic inputs. The Distributions file contains the SAS variables shown in Table F-2:

Table F-2. Column definitions for the distributions files.

Column Name	Format	Purpose
Ordering	num	used for sorting the records into the correct order
VarName	$24	Name of the input variable
Condition	$200	A SAS logical statement indicating when the distribution applies
Form	$16	the type of distribution (e.g. uniform, normal, beta,…)
v1	num	the first distributional parameter
v2	num	the second distributional parameter (if needed)
v3	num	the third distributional parameter (if needed)
Truncation	num	Flag: 0=no truncation, other values for various options
MinVal	num	Lower end truncation point (value or percentile)
MaxVal	num	Upper end truncation point (value or percentile)
ProbVector	$200	Comma-separated list of probabilities
Function	$200	SAS expression that defines the value of this input
EmpFile	$40	File name for empirical input data for this variable
UncFile	$40	File name for uncertainty data for this input
Notes	$200	Space for user notes

The VarName indicates which input is being described, and in each case it must match one of the VarNames on the Variables file. Multiple distributions may be defined for the same VarName, in which case conditions are needed to indicate when each distribution applies. Each condition amounts to a logical expression following the rules of SAS code, which the program evaluates as

IF (condition) THEN variable = …

When multiple distributions are used for one input variable, the list of conditions should cover all possible outcomes, or else it may happen that none of the IF statements are true and the variable remains undefined. This may result in missing values that could affect the results. If the condition is left blank, the same distribution is always used whenever this variable is evaluated.

The Form indicates the type of distribution. It is not case-sensitive, and only the first four characters of the form are used by SHEDS (so for example, the user may enter "expo" rather than "exponential"). The form may be set to any of the values indicated in Table F-3:

Table F-3. Distributional shapes

Form	Purpose
beta	a random value from a 2-parameter beta distribution
binomial	a random value from a binomial distribution (either zero or one)
empirical	a value randomly chosen from a list
exponential	a random value from a 2-parameter exponential distribution
function	a value defined by the specified function
gamma	a random value from a 2-parameter gamma distribution
lognormal	random value from a 2-parameter lognormal distribution
normal	random value from a 2-parameter normal distribution
point	a single point value
probvector	a random value chosen from a probability vector
triangle	a random value from a 3-parameter triangular distribution
uniform	a random value from a 2-parameter uniform distribution
Weibull	a random value from a 2-parameter Weibull distribution

The parameters v1, v2, and v3 have form-specific meanings as detailed in Appendix E. If Truncation=0 then the return value is never altered, even if its value is quite extreme. If Truncation=1 then values below MinVal are replaced by MinVal, while those above MaxVal are replaced by Maxval. It is possible to define one truncation point but not the other. Truncation=2 operates like Truncation=1, except that the truncation points are percentiles rather than return values. Truncation=3 uses percentiles but spreads any out-of-bounds values throughout the allowed region.

The ProbVector column is used only with probability vectors. The probability for each outcome is given in a comma-separated list. The return value is a position in the list, so the SHEDS code can only use probability vectors for variables in which this position has a pre-defined meaning. The code automatically adjusts the probabilities to sum to unity.

The Function column is not often used, but it is very flexible. The user may specify any SAS expression that refers to other variables currently known (meaning they have values defined on the same record of the same SAS data set as the input being defined). In principle, all the usual continuous distributions could be replaced by equivalent Function definitions. Furthermore, variables could be made to be explicit functions of one or more other variables. For example, the object-mouthing frequency could be set to a multiple of the hand-mouthing frequency. This is a powerful capability that can easily lead to errors, so the user must be very careful when using this option.

EmpFile contains the name of a SAS data set in the input\emp folder that is to be used as an empirical distribution for this input variable. This data set should have one column with the same name as the input being defined (for example, Dermal_Maxload if this input is the maximum dermal loading limit). One record is selected at random from this data set, and the value on that record is the new value for the input variable. Empfile is not used unless the Form=empirical.

UncFile contains the name of a SAS data set in the \input\unc folder that contains sets of distributional parameters that collectively define an uncertainty cloud. This data set may contain any or all of the variables listed in Table F-3 (except a new VarName, or a reference to another uncertainty file). As for an empirical variable, one record is selected from the UncFile at random and the values on that record replace the values currently being used. While this is often restricted to new values for v1, v2, and v3, it is possible to alter the Form, the truncation option or values, or the elements of a probability vector, and so on.

F.2.3 Agegroups file

The *Agegroups* file contains age-gender groups, with their status indicated by a Use variable. Use=1 means that the group is included in the current model run, while Use=0 indicates non-inclusion. The groupings are those that appear for the same purpose in the user interface, and are not necessarily those used for grouping activity diaries. In principle, the groups could be disaggregated, even down to single years of age, but this would be difficult to view via the interface. It would also require a lot of boxes to be checked to include several decades of ages. For batch runs of SHEDS, this is not a problem, and customized Agegroups files could be prepared.

F.2.4 AreaTreated File

The *AreaTreated* file contains one record for each indoor application scenario, with the same variables as the Distributions file. The purpose of this file is to define the time fractions spent in treated indoor areas. If a distribution is specified, then each person will sample the distribution once. A point value may be used to assign the same fraction to all persons. The variables on this file are used in the determination of contact time; see sections 2.6.2.and 2.6.3 for more details.

F.2.5 Correlations file

The *Correlations* file is used for specifying the Spearman correlations that are desired between pairs of continuous input variables. All variables listed on the Variables file, except those of the Product class, may be correlated. The correlations file has three columns or SAS variables shown in Table F-:

Table F-4. Column definitions for the correlation files.

Variable	Format	Purpose
var1	$24	name of first input variable in pair
var2	$24	name of second input variable in pair
corr	num	desired Spearman correlation

The order of the two variables in the pair does not matter, and only needs to be specified once (the reverse order is not required). The user may define as many pairs as they wish. Any correlations not explicitly listed on this file are assumed to be zero. This file may be empty if no correlations are requested. It is also possible to leave this file as is, while turning off the switch that implements the correlation algorithm. In that case, any data on this file is ignored.

Certain combinations of correlations are incompatible. Essentially, the correlation matrix must be positive definite (that is, it has no negative eigenvalues) for a solution to be possible. The model

constructs the matrix from the list on this input file, and reports invalid correlation matrices on the SAS log.

F.2.6 Dates_Fixed file

The Dates_Fixed file is used only if the user chooses to supply the usage dates for each application scenario. the alternatives are to use Dates_Variable or None; in the latter case the user supplies chemical concentration time series directly. The Dates_Fixed file contains one record for each application scenario, as shown in Table F-5:

Table F-5. Column definitions for the Dates_Fixed file.

Variable	Format	Purpose
Product	$12	application scenario name
Usage_Dates	$500	comma-separated list of day numbers
Usage_hour	num	hour of the day when application occurs (range 1-24)
Reentry_Hours	num	number of hours after application when contact is prohibited
Handlers	num	1=users over minimum age are handlers, 0=no handlers

The application dates are indicated by number, with day 1 being the first day of the simulation period. Zero or negative numbers may be entered, which would result in an application before the start of the activity diary and the determination of exposure. Such applications may cause concentrations that persist into the modeling period. SHEDS does not request the user to enter chemical concentrations at the start of the simulation period (unless the TimeSeries option is used), but instead calculates concentrations based on chemical usage patterns. For example, usage on day -6 would mean that the chemical was applied one week before the first day of the exposure simulation period. Usage days can be listed beyond the end of the simulation period, but these have no effect. For example, usage might be specified on day 180 of a year-long simulation. If the simulation length were then shortened to 90 days, the Dates_Fixed file would not have to be altered. Of course, if all the usage dates are beyond the end of the simulation period, then no exposure (other than background) can occur.

The Usage_hour is a number between 1 and 24 inclusive, with 1 indicating midnight and 24 indicating 11 p.m. This is slightly counterintuitive, as a range of 0-23 would seem more natural. The Reentry hours is the number of hours following application during which time in the treated area is set to zero, regardless of the activity diary. The reentry time may be set to zero, which means there is no such ban. Note that contact does not necessarily begin once this period is over. It is also necessary that the activity diary place the person in the right micro for contact to occur. For example, an indoor treatment at 9 a.m. may have a two-hour ban, but exposure to a child might not occur until the diary brings them home from school at 4 p.m.

The Handlers status flag is application scenario-specific. Under Fixed Dates, all persons over the handler age limit are designated as handlers, or else none are. For example, the placing of ant traps is often done by the homeowner, but termite treatments are usually by professionals. If the user would some some persons to be randomly designated as handlers while others are not (for the same application scenario), then use the Model-determined dates option.

F.2.7 Dates_Variable file

The *Dates_Variable* is required only if the user chooses to have the model randomly assign chemical usage patterns. In that case, each simulated person will have their own pattern, and some people might not use any of the application scenarios at all. For each application scenario the user supplies the

probability of usage, a vector of the usage frequency, probability vectors for month and day of week on which usage occurs, and optional co-occurrence rules if the usages are not independent. Table F-6 indicates the structure of the Dates_Variable file.

Table F-6. Column definitions for the Dates_Variable file.

Variable	Format	Purpose
Product	$12	application scenario name
requirement	$40	logical condition necessary for application to occur
p_user	num	probability of this application scenario occurring at this house (in one year)
p_handler	num	probability that simulated person is the handler (if they are of age)
v_NumApps	$500	probability vector for usages per year (starting at one usage)
v_Month	$200	probability vector for month of usage
v_Weekday	$100	probability vector for day of week
blackout	num	minimum number of days between multiple uses of the same scenario
ProbMultiplier	num	multiplicative factor for probabilities within co-occurrence window
Ndays	num	number of days on each side of usage in co-occurrence window
Co_occur_list	$60	list of scenario numbers that create usage windows

In a single-application scenario model run, p_User is usually set to one so that only users are modeled. Non-users will have no exposure except possibly for outdoor background. In a multiple scenario run, each p_user should be set to the appropriate level to ensure the correct frequency of the same person using more than one application scenario. This will have the side result that some simulated persons will use no application scenarios at all, but cannot be avoided. Note that p_user applies only to those persons who meet any requirement for the application scenario. For example, for pet application scenarios, p_user applies only to persons who have one or more suitable pets. For instance, if 50% of the population have a pet and p_user=20%, then only 10% of the simulated people will use this pet application scenario.

Persons below the minimum age for handlers will never be handlers in SHEDS, regardless of the value of p_handler. That value applies only to persons at or above the minimum age. Even if a given application scenario is exclusively used by homeowners (as opposed to professional application), p_handler may be less than one because there may be multiple adults per house and typically, only one person handles the chemical. Of course, the user may wish to confine the population to handlers only, in which case p_handler should be one.

The probabilities in the v_NumApps vector sum to one, which implies that the person has already been designated as a user. The probability of zero uses is not included since that may be conditional on other variables. For example, lawn, garden, or pet application scenarios can only have uses if the simulated person happens to have a lawn, garden, or pet, respectively. If the general population is being modeled, then many people will not qualify to be users of particular application scenarios. Changing the fraction of the population who meet these requirements may be made without changing the probabilities in the v_NumApps vector. Note that v_NumApps refer to the number of usages of this application scenario in a one-year period, regardless of the length of the SHEDS simulation. Thus, the user may alter the simulation period without having to alter the v_NumApps data.

The v_Month and v_Weekday vectors are long-term averages covering a full calendar year. Thus, the first element in v_Month is the fraction of all applications of this application scenario that occur in January, regardless of the simulation period for the current model run. SHEDS was set up this way to simplify the input data, as discussed above for v_NumApps.

The blackout period refers to the minimum number of days between multiple applications of the same application scenario. For example, if blackout=7 then two applications cannot occur less than one week apart. The code imposes a minimum blackout period of one day, which means that two applications of the same application scenario cannot happen on the same day.

The default assumption in SHEDS is that the usage dates for each application scenario are assigned independently. The final three variables are used only if the user wants to alter the usage probabilities for one application scenario based on the usage dates assigned to other application scenarios. For example, if the same product can be used on both lawns and gardens, then the dates for a garden application scenario may tend to coincide with those assigned to a lawn application scenario. In that case, the lawn dates are chosen first, meaning that all the lawn application scenario usage dates are chosen throughout the simulation period, before any garden dates are assigned. Then the probabilities for the possible garden dates will be modified by their proximity to the selected lawn dates. The Ndays variable refers to the number of days in the *window* around each lawn usage date. A window width of N affects a total of (2N-1) days around each usage date. For example, Ndays=2 means the day before, during, or following a lawn usage would have its probability for a garden usage altered. Ndays=1 would only alter the probabilities on the actual lawn usage dates. The probabilities on the dates inside the window are each multiplied by the ProbMultiplier value. If this is greater than one then the dates inside the window have enhanced probability, whereas if ProbMultiplier is less than one then these dates have diminished probability. If the original probability was zero (for example, perhaps a garden treatment cannot happen in December), then the altered probabilities will remain zero, regardless of the co-occurrence settings.

The co_occur_list may include one or more application scenarios. For example, suppose application scenario #3 is a lawn application scenario and #6 is a garden application scenario. Then the co_occur_list for application scenario six might simply list the number 3. However, if application scenario #4 were another lawn application scenario and it also should affect #6, then "3,4" should be listed instead. A blank co_occur_list automatically means that the dates for the given application scenario are selected independently of the dates for other application scenarios. However, it would still be possible for a later application scenario to depend on the current one. For example, even if the dates for application scenario #6 are selected independently of the first five, application scenario #7 (or later) could depend on the dates chosen for #6. SHEDS does not iterate the date selection; each application scenario is processed once (per person) in the application scenario priority order, as given on the Scenarios input file.

F.2.8 Decay file

The Decay file is used only if the decay-dispersion option is selected for chemical concentrations. (The alternatives are the Intervals or TimeSeries options). Each application scenario has multiple records on the Decay file, with the number of records depending on the scenario location. A pet scenario has only two records, one for the surface concentration on the pet immediately after application and the other for the fraction lost per day (sometimes called the *decay rate*, although technically the latter term more properly refers to the rate constant in an exponential decay equation). For small values the two are very close, but while the fraction lost per day cannot exceed one, the decay rate can be any positive quantity. For example, if the fraction lost per day is 5% then the true decay rate is 0.051 day^{-1}, which is not appreciably different given the measurement error. But if the fraction lost per day were 90%, then the true decay rate constant would be 2.30 day^{-1}. But even when the phrase "decay rate" is used in SHEDS, it refers to the fraction lost per day.

A lawn or garden application scenario has three records on the Decay file, one for chemical surface loading immediately after application, one for soil concentration, and the final one for the fraction lost per day. The surface loading is expressed in micrograms per square centimeter, and the user should

decide whether this represents the total amount of chemical present or just the amount available for transfer. For example, the transfer efficiency approach usually refers to the former, while the transfer coefficient approach refers to the latter. The soil concentration is expressed in micrograms of chemical per gram of soil. SHEDS converts the soil concentration to an equivalent surface loading and then adds it to any surface loading specified directly. Therefore, the user should specify both a surface loading and a soil concentration only if there is no danger of "double counting" the same chemical twice. If the surface loading represents the chemical application rate, then no soil concentration should be specified.

An indoor scenario has six records on the Decay file. Two are for the surface loading on hard floors and carpets, respectively. The third is for the concentration in dust. A fourth is for the air concentration. The fifth record is for the fraction lost per day. In SHEDS version 4 this fraction is common to all affected media, because the fugacity model being developed for SHEDS (Glen et al., 2007) has shown that the medium that decays the slowest acts as a reservoir that replenishes the other media, so that they tend to track its concentration profile over time. The sixth input for each indoor scenario gives the ratio of surface loading in the untreated parts of the house to the surface loading in a treated area, at the time the former reaches its peak. This determines the rate at which the chemical disperses into the untreated areas. For high volatility chemicals this ratio would be relatively large (say, in the range 0.1 to 1.0). However, most household pesticides have low volatility because the natural loss rate for airborne chemical is extremely high due to air exchange with the outdoors. With low volatility chemical the rate of dispersion into untreated areas is low, usually leading to a ratio below 0.01. For ratios below this value, the untreated areas contribute very little to the overall exposure. Table F-7 shows the SAS variables present in the Decay input file.

Table F-7. SAS variables in the Decay input file.

Column name	Format	Purpose
Priority	num	used to sort the records (if necessary)
Chemical	$32	name of chemical agent
Product	$12	scenario name
VarName	$24	name of input variable
Units	$16	units of input variable
Form	$12	distribution type (e.g. uniform, normal,…)
v1	num	the first distributional parameter
v2	num	the second distributional parameter (if needed)
v3	num	the third distributional parameter (if needed)
truncation	num	flag for type of truncation (0=no truncation)
minval	num	the lower truncation point
maxval	num	the upper truncation point
uvar	$26	name of the u_variable that determines this input
description	$80	description of this input variable
probvector	$200	probability vector values for this input
function	$200	SAS expression for this input
empfile	$40	name of SAS data set with empirical data
uncfile	$40	name of SAS data set with uncertainty data
notes	$200	user documentation

If the model run contains multiple chemicals, then the records described above are repeated for each chemical. Thus, with 3 chemicals there would be 6 records for each pet application scenario, 9 for lawn or garden scenarios, and 18 for indoor scenarios. The records on the decay file are appended to

those on the Distributions and similar files, so any of the features described there (such as the empirical data option, functions, or uncertainty) may be used.

The input variables have the following names

VarName	Description
avail_THard	Surface loading after application on indoor treated hard floors
avail_TSoft	Surface loading after application on indoor treated carpets
avail_TDust	Chemical concentration in dust in treated indoor areas
avail_TAir	Air concentration after application in treated areas
avail_Lawn	Surface loading after application on lawns
avail _LSoil	Chemical concentration in soil after lawn application
avail_Garden	Surface loading after application in vegetable gardens
avail_GSoil	Chemical concentration in soil after garden application
avail_Pet	Surface loading on pet immediately after application
f_Decay	Scenario-specific fraction of chemical lost per day
UT_ratio	Untreated/treated ratio at peak of untreated concentration

F.2.9 Intervals file

The *Intervals* file is needed only if the Intervals method is used to determine chemical concentrations. This method requires four distributions for each input variable, for each application scenario and chemical. The list of input variables is the same as for the Decay method, except that the fraction lost per day is not required. Hence, per chemical, there are four distributions for each pet scenario, eight for each lawn or garden application scenario, and 20 for each indoor scenario. However, any inputs that are always zero may be omitted from the input file, which may reduce the numbers.

The four distributions for each input correspond to four time periods, measured from the most recent application of the given application scenario. These periods are

Period	Description
0	Day of application (day 0), from application time until midnight
1	From 1 to 7 days inclusive after application
8	From 8 to 30 days inclusive after application
31	31 or more days since the most recent application

The period appears as a numerical suffix on variable names in the SAS code. The basic variable names (without the suffix) are the same as those listed as VarNames under the Decay option, except that f_Decay is not used.

F.2.10 Scenarios file

The *Scenarios* file lists the possible scenarios for inclusion in the model run. Unless new scenarios are defined, the only column that the user may need to change is the Use variable, which is set to 1 to include a scenario, or 0 to exclude it. The columns on this file are shown in Table F-8.

Table F-8. Column definitions for the Scenarios file.

Variable	Format	Purpose
priority	num	for sorting the scenarios (usage dates are assigned in this order)
product	$12	scenario name
use	num	flag (1=include, 0=exclude from model run)
description	$80	scenario description
location	$8	scenario location

The scenario locations are important if new application scenarios are defined. The four locations are lawn, garden, pet, and indoor. The location determines which requirement applies (Has_lawn, Has_garden, Has_pet, or none), and also determines which contact media may be affected. When fewer scenarios than are on the default file are run, the most convenient option is to set Use=0. Instead, it is possible to delete those records from the Scenarios file completely, but that makes it more difficult to recover the later for another
run.

F.2.11 TimeSeries file

The *TimeSeries* file is needed only if the user chooses to supply chemical concentration time series directly, rather than having them calculated by SHEDS from the usage patterns. In SHEDS version 4 the TimeSeries input file differs from the one in version 3 by the need to supply the chemical name on each record.

The TimeSeries input file is discussed in some detail in Chapter 2, with an example file provided in Table 2-3. The file contains one record per day. The required variables are the chemical name and the date (in SAS date format). The file also contains one or more media names from the list (THard, TSoft, Tair, Lawn, Garden, and Pet), followed by an underscore and a suffix representing the house number. The house number allows the user to supply multiple examples for each medium. For each simulated person a house number is selected at random and all the time series from that house are used together. It is expected that the media within houses will be correlated with each other, but not necessarily across houses.

Any media omitted from the TimeSeries file are assumed to have zero concentration at all times. In the example in Table 2-3 only THard and TSoft appear, so TAir, Lawn, Garden, and Pet are assumed to be zero. Different media may be specified in each house. For example, some houses may have pets whiles others do not. Each time series should be free from internal gaps (missing days), and all the media belonging to a given house should start and stop on the same date. However, different houses may start and stop on different dates, and do not have to overlap at all. In practical terms, this form of input is most often used in conjunction with a field study in which the same set of monitors may be moved from one house to another each week or so.

When using the TimeSeries option, the chemical concentration is not tied to specific scenarios and there is no need to specify usage data. Also, since the treated area in each house is scenario-specific,

it is not defined under this option. Therefore, the untreated media are not used, and the "treated" media should represent an average of all measurements at the given house. In effect, it is a weighted average of treated and untreated values which is encountered 100% of the time that the simulated person is in the appropriate micro.

At present, the TimeSeries option can handle daily measurements only. SHEDS internally converts these to hourly time series by repeating the same concentrations 24 times each. No interpolation is used. If there is a need, it would be possible for future versions of SHEDS to accommodate hourly input data.

F.3 Brief Description of Other Input Files

F.3.1 Bwsa file

This file contains distributions for body weight, height, and basal metabolic rate, as functions of age and gender. There are 200 records, one for each year of age (0-99) for females, and then the same for males. Below age 20 years it provides regression parameters (mean and standard deviation) for height and the logarithm of weight, for each age and gender, along with the slope, intercept, and magnitude of the residuals for the basal metabolic rate as a function of weight. For age 20 years and above, the logarithm of weight is selected from a normal distribution as for younger persons, but now the height is determined from body weight rather than from age. The slope, intercept, and residual for height as a function of log(weight) are provided. The basal metabolic rates are determined from regression parameters, just as for ages below 20 years.

F.3.2 Cohorts file

This file is used to define the pools for the activity diaries. It divides the population into groups by gender and age range, and further divides the simulation days by season and daytype. The current example of this file contains 14 age-gender cohorts and 8 season-daytype combinations, for a total of 112 records. Each age group is defined as a range between a specified minimum age and maximum age. If the user desired, the number of age cohorts for diary assembly purposes could be changed by suitably altering this file.

F.3.3 ContactMedia file

The C*ontactMedia* file contains a list of the potential contact media in SHEDS. The current version of the model has 9 contact media, listed in Table 2-1. The SAS variables on this file are shown in Table F-9:

Table F-9. Column definitions for the contact media file.

Variable	Format	Purpose
medium	$6	name of contact medium
ShortDescription	$24	brief description of contact medium (for use in tables)
LongDescription	$60	longer description of contact medium

The list of contact media in SHEDS version 3 was flexible. In version 4 this list is fixed, which makes the code more readable. Contact media can effectively be deleted by assigning zero concentration to them. Adding media would require code changes in version 4.

F.3.4 Diaryevents file

This is the large file (102 megabytes) containing all the diary events from all persons in the CHAD database. It is only of interest to the user if they wish to add to it or replace it with different diary data. The Diaryevents file is created by the Diary PreProcessing module in SHEDS.

Table F-10. Column definitions for the diaryevents file.

Variable	Format	Description
chadid	$9	CHAD personal ID (uniquely identifies the diary day)
start	num	event start time in minutes past midnight (0-1439)
duration	num	event duration in minutes
act	$5	CHAD activity code
loc	$5	CHAD location code
special	$1	"Z"=sleeping event, "B"=bath/shower event
agegroup	num	cohort number
micro	$8	SHEDS microenvironment
io	$1	"I"=indoors, "O"=outdoors
metsdist	num	METS distribution number
meanmets	num	average METS for this activity

The version of the Diaryevents file in the SHEDS default directory contains 1,441,185 records (diary events), covering 33,748 diary days.

F.3.5 Diarylocs file

This file contains the mapping from CHAD location codes to the microenvironments used in SHEDS. There are only three SAS variables: the CHAD location code, a brief description of the CHAD code, and a SHEDS microenvironment (micro).

Table F-11. Current microenvironments used by SHEDS.

Micro	Description
in_home	inside a residence
out_home	outside near a residence
in_oth	inside another building
out_oth	outside away from residence
in_veh	in a vehicle

F.3.6 DiaryQArules file

This file contains a list of logical rules for accepting a CHAD diary for use in SHEDS version 4. For instance, one might keep diaries from a limited range of ages, or perhaps those with no outdoor time. The list of rules on the demonstration input file is:

- (gender) in ('M','F')
- (age>=0) and (age<=99)
- LocUX<=240

- ActUX<=360
- Pai < 4
- Qceattime<=360
- Qcactloc<=360

The first two of these rules are self-explanatory. The last five refer to the total time spent on the given diary day in various conditions. Rules 3 and 4 put limits on the amount of time in locations U and X (this means location is unknown or missing), and the amount of time in activities U and X (which means the activity code and hence the METS level are missing). The fifth rule limits the physical activity index (which is the duration-weighted average METS value) to be less than 4. Only a few diaries exceed this, and those are suspected to have coding issues. Similarly, the last two rules are designed to eliminate a small number of diaries with an unlikely activity sequence, probably indicative of coding problems.

The list of rules can be modified. Each record contains one logical test that is applied to the diary-day as a whole. Partial days are never kept or used. All tests must pass for a diary to be retained for later use.

F.3.7 Diaryquest file

This file contains the personal data from CHAD that do not change over the course of the diary day, like age, gender, home state, diary date, day of week, season, temperature, and many other variables. This file is used to determine if the diary passes the QA tests on the diaryQArules file, and for assigning the CHAD diaries to appropriate cohorts and pools. A list of the variables on this file is given in the Activity Diary-related Inputs section of the Model Inputs chapter.

F.3.8 MetsDists file

The *MetsDists* file contains the METS distribution for each activity in CHAD. There are 91 distributions on the current MetsDists file, ordered by decreasing frequency of occurrence in the CHAD database. (This ordering is used to speed up the SHEDS code slightly by only searching the distribution list until a hit is found.)

F.3.9 MetsMap file

The *MetsMap* file lists the CHAD activity code and the corresponding distribution number on the MetsDists file. Some of the activities for sports or exercise have age-dependent distributions, so age is used to obtain the mapping in these cases.

F.3.10 Pop2000 file

This file contains a summary of the 2000 U.S. Census, without any geographical detail. It contains counts by age, gender, and race combination, and summaries across ages and across race. It is used to select the age and gender of the randomly generated simulated individuals.

APPENDIX G Default Values for Non-chemical Specific Variables

Scenario	SHEDS Variable	Units	Model Input	Notes	Reference
General	Dust loading on carpet	g/m²	Lognormal (7.8, 2.9)	Based on N=376 for vacuum samples from carpets.	• Adgate et al. (1995)
	Dust loading on hard floors	g/m²	Lognormal (0.42, 2.8)	Based on dust loading measurements for samples collected in 216 homes (N=444).	• Adgate et al. (1995)
	Body-surface fractional contact rate	fraction_of_skin_surface/hour	Beta (20.2, 13.1)	Assume total 3 hours per day. Distributions fit to hand-to-playset contact information from videotapes of four 5-7 year-old children and hand-to-hard floor contacts/hr from a 4-child study.	• Freeman N; personal communication
	Hand-surface fractional contact rate	fraction_of_skin_surface/hour	Weibull (9.1, 2.5)	Assume total 3 hours per day. Distributions fit to hand-to-playset contact information from videotapes of four 5-7 year-old children and hand-to-hard floor contacts/hr from a 4-child study.	• Freeman N; personal communication
	Dermal brush-off rate	fraction/hour	0	Assumption	
	Fraction of dermal loading removed during bath/shower		Uniform (0.9, 1.0)	Assumption. This represents the fraction of dermal soil/dust loading removed from the hands during a bath or shower. No suitable data sources have been identified yet. Bathing is assumed to be very efficient for removal.	
	Fraction of hand loading removed during hand washing		Uniform (0.3, 0.9)	Assumption. Like bathing removal efficiency, no good data sources have been identified. Hand washing is assumed to be less efficient and more variable than bathing, based on its shorter duration. However, hand washing occurs much more frequently than bathing, so it has more potential to reduce hand-to-mouth transfer.	

	Fraction of hand surface area that enters mouth		Beta (3.7, 25)	For hand size 200 cm2, this translates to a mean area of about 26 cm2, or about 1 ½ fingers for a child.	• Zartarian et al., 2005; based on Leckie et al., 2000
	Fraction of TC and Handler exposure going to hands		0.2	Assumption. Dermal transfer method: TC only variable Handler only variable	
	Hand mouthing frequency - indoors	1/hour	Weibull (0.73, 11.96) max = 60	Modified from Xue 2007	• Xue et al., 2007
	Hand mouthing frequency - outdoors	1/hour	Weibull (0.55, 5.53) max = 60	Modified from Xue 2007	• Xue et al., 2007
	Hand-to-mouth transfer efficiency (saliva removal efficiency)		Beta (2.0,8.0)	Consistent with literature values after accounting for study protocols. We assume a Beta(2,8) distribution which has a 5th percentile near 0.05 and a 95th percentile near 0.52, with a mean of 0.20.	• Kissel et al., 1998 • Nishioka et al., 2003
	Maximum dermal loading limit	µg/cm²	Triangle (0.1, 1, 2.1)	Hand loading	• Cohen Hubal et al., 2005
	Object-to-mouth contact surface area	cm²	Exp (1, 10) Max=50	Based on the area of hand mouthed by 2-5 year olds from Leckie 2000 and the assumption that children mouth less area of objects than their hand. The maximum is comparable to the surface area of a ping-pong ball.	• Leckie et al., 2000
	Object-to-mouth transfer efficiency (saliva removal efficiency)		Beta (2.0, 8.0)	Assume same as hand-to-mouth transfer efficiency.	

	Object-mouth contact frequency - indoors	1/hour	Weibull (0.6, 6.8) max = 60	Modified from Xue 2010	• Xue et al., 2010a
	Object-mouth contact frequency - outdoors	1/hour	Weibull (0.55, 5.38) max = 60	Modified from Xue 2010	• Xue et al., 2010a
	Soil ingestion rate (mean daily)	mg/day	Lognormal (17, 4) min = 1 max= 257	Assumed lognormal with mean and std. dev. Reported by Ozkaynak 2010 with min and max at 2.5 and 97.5 percentiles.	• Ozkaynak et al., 2010
	Fraction of body unclothed		Beta (3, 6.7)	Warm weather conditions.	• Zartarian et al., 2005; references therein - Wong et al. 2000; EPA CSEFH (2008)
	Max # days between baths	days	Multinomial [.75, .14, .07, .01, .01, .01, .01]	Warm weather conditions. The multinomial elements correspond to [1, 2, 3, …, 7 days]	• Zartarian et al,, 2005 - referencing Kissel (2003, personal communication; raw SCS-II data for warm weather) • Shoaf et al., 2005

	Mean # hand washes per day per person		Lognormal (3.6, 2) Min=1 Max=12	This is a continuous variable used to set the probability of hand washing. The number of actual hand washes varies from day to day, as do the specific times when the washing occurs.	• Zartarian et al,, 2005; references therein -- Tsang and Klepeis 1996; Freeman et al., 2001; and Kissel (2003, personal communication; raw SCS-II data for warm weather) • Shoaf et al., 2005
Indoor Scenarios Only	Size of house	m^2	Lognormal (168,1.6) Min=20 Max=1000	Handler only variable	• American Housing Survey (U.S. Department of Housing and Urban Development and U.S. Census Bureau), 2007
	Fraction of in-home time spent on carpet		Beta (4.44, 4.12)	Based on the means and std dev from 7:30am-7:30pm for ages from 22 - 31 months.	• Brinkman et al., 1999
Veg. Garden Scenarios Only	Fraction of outdoor time at home spent in vegetable garden		Uniform (0 , 0.1)	Assumption based on NGA report.	• National Gardening Association, 2009
	Pr (having a vegetable garden)		0.23		• National Gardening Association, 2009
	Size of vegetable garden	m^2	Lognormal (9, 6.78) Min=1 Max=1000	Handlers only.	• National Gardening Association, 2009

Lawn Scenarios	Fraction of outdoor time at home spent on lawn		Beta (6, 2)	Beta assumed with mean of 75%.	
	Pr (having a lawn)		0.47		• Johnson et al. 1999
	Size of lawn	m^2	Uniform (46, 1393)	Handlers only.	• US EPA 2006a
Pet Products	Fraction of home time spent near pet		Exp (0.0044, 0.0185) Max=0.1337		• Freeman et al., 2001 (op cit. OPP) combined with CHAD
	Pr (having a dog or cat)		0.58	0.696 - 0.372 * 0.324 = 0.575 (Pr. having a dog plus Pr. having a cat) - (Pr. having a dog)*(Pr. having a cat)	• American Veterinary Medical Association, 2007
	Size of pet	m^2	Cats (0.15, 0.25, 0.40) Dogs (0.30, 0.70, 1.10)	Handlers only. Values correspond to small, medium and large sized animals.	• US EPA 2009, from Stahl, 1967 op. cit. EPA Wildlife Exposure Fact. Hand., 1993