US EPA ARCHIVE DOCUMENT

United States Environmental Protection Agency

Stochastic Human Exposure and Dose Simulation Model for Multimedia, Multipathway Chemicals

SHEDS-Multimedia Model
Version 3
Technical Manual

US EPA ARCHIVE DOCUMENT

US EPA ARCHIVE DOCUMENT

EPA 600/R-08/118
September 2008
www.epa.gov

Stochastic Human Exposure and Dose Simulation Model for Multimedia, Multipathway Chemicals

SHEDS-Multimedia Model
Version 3
Technical Manual

Valerie Zartarian[1]
Graham Glen[2]
Luther Smith[2]
Jianping Xue[1]

[1]US Environmental Protection Agency
National Exposure Research Laboratory

[2]Alion Science and Technology, Inc.

U.S. Environmental Protection Agency
Office of Research and Development
Washington, DC 20460

US EPA ARCHIVE DOCUMENT

Acknowledgments

We would like to thank Dana Vogel, Steve Nako, David Hrdy, and Phillip Villanueva of the U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxic Substances, Office of Pesticide Programs and Miles Okino of the U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory for providing comments on the draft Technical Manual.

US EPA ARCHIVE DOCUMENT

TABLE OF CONTENTS

Acknowledgments i
TABLE OF CONTENTS ii
LIST OF FIGURES vi
LIST OF TABLES vii
ACRONYMS AND ABBREVIATIONS ix
1 INTRODUCTION 1
1.1 Purpose of SHEDS-Multimedia 1
1.2 SHEDS-Multimedia Model Development History 2
1.3 SHEDS-Multimedia Version 3 Overview 3
1.4 Features and Advantages of SHEDS-Multimedia Version 3 7
2 MODEL STRUCTURE AND APPROACH 9
2.1 Defining Exposure Scenarios and Simulation Information 9
2.2 Creating a Simulated Individual 10
2.2.1 Age Groups 10
2.2.2 Assigning ages to individuals 11
2.2.3 Assigning other person-level variables 11
2.3 Generating Longitudinal Activity Patterns for a Simulated Individual 12
2.3.1 Selecting Activity Diaries 12
2.4 Simulating an Individual's Contact with Scenario-Relevant Media 13
2.4.1 Contact Media 13
2.4.2 Determination of contact events 14
2.5 Generating Concentration Time Series for the Contact Media 16
2.5.1 Chemical usage patterns (applications) 16
2.5.2 Media concentration generation methods 22
2.6 Generating Exposure Time Series for Individuals 26
2.6.1 SHEDS equations for exposure by pathway 26
2.6.2 New Exposure 28
2.6.3 SHEDS Exposure Adjustments 33
2.7 Calculation of Dose Time Series for the Individual 36
2.8 Extract Daily Statistics from Exposure or Dose Time Series 37
2.9 Generating Population Variability Estimates of Exposure or Dose 37

US EPA ARCHIVE DOCUMENT

3 MODEL INPUTS 38
3.1 Stochastic Sampling 39
3.1.1 Probability vectors 40
3.1.2 Continuous Probability Distributions 41
3.1.3 Sampling frequency 44
3.1.4 Correlation of Inputs 44
3.2 General Issues 45
3.2.1 Units 45
3.2.2 Formats 45
3.2.3 Demo File and multiple distributions for one variable 46
3.2.4 Simulation period 46
3.3 Chemical Usage-Related Inputs 47
3.3.1 Application-specific inputs 47
3.3.2 Co-Occurrence Factors 49
3.4 Activity Diary-Related Inputs 49
3.4.1 CHAD Diaries 49
3.4.2 Re-extracting data from CHAD for use in SHEDS 50
3.4.3 Pre-processing the CHAD files 50
3.5 Contact Probability-related Inputs 54
3.6 Concentration-Related Inputs 55
3.6.1 Decay-dispersion method 55
3.6.2 Post-application method 56
3.6.3 Time Series method 56
3.7 Exposure-Related Inputs 57
3.7.1 Physical exposure factors 57
3.7.2 Activity-related exposure factors 58
3.7.3 Transfer-related exposure factors 60
3.8 Dose Factors 61
4 OUTPUTS 63
4.1 Individual Time Series 63
4.2 Population Variability Distributions 63
5 SENSITIVITY ANALYSES 66
5.1 Pearson Correlation 67
5.2 Spearman Correlation 67
5.3 Stepwise Regression 68

US EPA ARCHIVE DOCUMENT

5.4 The Percentile Scaling Approach 70
6 UNCERTAINTY ANALYSES 71
6.1 The Bootstrap Approach for Input Uncertainty Distributions 72
6.2 Uncertainty inputs and outputs 73
6.3 Sensitivity Analysis on Results from a Two-Stage Monte Carlo Model Run 75
7 QUALITY ASSURANCE (QA) 78
7.1 Development QA 78
7.2 One-person, one-day simulation for code verification 78
8 REFERENCES 80
APPENDIX A Terminology Used in SHEDS-Multimedia A-1
APPENDIX B Selected SAS Variable Nomenclature B-1
APPENDIX C The SHEDS Code Modules and their Functions C-3
C.1 List of SAS Code Modules and Their Calling Sequence C-6
APPENDIX D Summary of the CHAD Database D-1
APPENDIX E Probability Density Functions E-1
E.1 Beta E-1
E.2 Exponential E-1
E.3 Gamma E-1
E.4 Lognormal E-2
E.5 Normal E-2
E.6 Point E-2
E.7 Triangular E-3
E.8 Uniform E-3
E.9 Weibull E-3
E.10 Discrete Probability Density Functions E-3
APPENDIX F Description of SHEDS Multimedia Input Files F-1
F.1 List of Input Files F-1
F.2 Detailed Description of User-Specified Input Files F-2
F.2.1 Variables file F-2
F.2.2 Distributions file F-4
F.2.3 Correlations file F-6
F.2.4 Contactmedia file F-6
F.2.5 ModelDates file F-7
F.2.6 UserDates file F-7
F.2.7 AppMedia file F-8

US EPA ARCHIVE DOCUMENT

F.2.8 Diarydetails file F-9
F.3 Brief Description of Other Input Files F-10
F.3.1 bwsa file F-10
F.3.2 diarycohorts file F-10
F.3.3 diaryevents file F-10
F.3.4 diarylevels file F-10
F.3.5 diarylocs file F-11
F.3.6 diarymets file F-11
F.3.7 diaryQArules file F-11
F.3.8 diaryquest file F-12
F.3.9 pop2000 file F-12
APPENDIX G Tables for the One-Person, One-Day Simulation for Code Verification G-1

US EPA ARCHIVE DOCUMENT

LIST OF FIGURES

Figure 1-1. Hypothetical exposure profile for an individual.....4

Figure 1-2. Monte Carlo sampling in SHEDS. 5

Figure 1-3. Overview of SHEDS methodology. 7

Figure 2-1. SHEDS approach for simulating one-year activity patterns.....13

Figure 2-2. Choosing application dates for a single treatment type.....19

Figure 2-3. Co-occurrence of treatment types, choosing application dates for treatment type 2.....20

Figure 2-4. Example time profile of dermal exposure from surface residues.....26

Figure 2-5.Example time profile of GI tract exposure from surface residues.....27

Figure 4-1. Example cumulative density function (CDF) plot from a SHEDS variability run.....64

Figure 6-1. Hypothetical uncertainty analysis results: CDFs for low-, medium-, and high-dose populations, as ranked by their medians. 74

Figure 6-2. Hypothetical uncertainty analysis results: CDFs for the 5th, 50th, and 95th percentiles from all simulated populations. 74

Figure C-1. Conceptual overview of SHEDS-Multimedia version 3 module linkages. C-4

US EPA ARCHIVE DOCUMENT

LIST OF TABLES

Table 2-1. Default list of contact media in SHEDS-multimedia version 3. 14
Table 2-2. Micro environments (micros) in SHEDS. 15
Table 2-3. Chemical application types in SHEDS. 17
Table 2-4. Example input file for concentration time series. 25
Table 2-5. Exposure pathways in SHEDS. 27
Table 3-1. Summary of SHEDS user-specified input variables (not all apply to all model runs). 39
Table 3-2. List of user-supplied probability vectors. 41
Table 3-3. SHEDS input probability distribution types. 42
Table 3-4. List of continuous variables in SHEDS. 43
Table 3-5. Rules for diary pre-processing in SHEDS. 51
Table 3-6. Variables that may be used in diary QA rules. 51
Table 3-7. Contact probability variables. 54
Table 4-1. Output files for SHEDS-multimedia version 3 variability run 65
Table 5-1. Hypothetical stepwise regression and percentile scaling sensitivity analysis results. 69
Table 6-1. Hypothetical Spearman correlation uncertainty analysis results. 76
Table 6-2. Hypothetical stepwise regression uncertainty analysis results. 77
Table D-1. CHAD Studies used in SHEDS. D-1
Table D-2. Diary Counts for SHEDS Age-Gender Cohorts. D-1
Table D-3. Example portion of CHAD Diary. D-2
Table D-4. CHAD Activity Codes. D-2
Table D-5. CHAD Location Codes. D-6
Table F-1. Column definitions for the variables file. F-3
Table F-2. Definitions for variables from the distributions and variables files. F-3
Table F-3. Column definitions for the distributions file. F-4
Table F-4. Column definitions for the correlations file. F-6
Table F-5. Column definitions for the contactmedia file. F-6
Table F-6. Default contact media in the contactmedia file. F-7
Table F-7. Column definitions for the modeldates file. F-7
Table F-8. Column definitions for the userdates file. F-8
Table F-9. SAS variables in both appmedia files. F-8
Table F-10. Contact probability definitions. F-9
Table F-11. Column definitions for the diaryevents file. F-10

US EPA ARCHIVE DOCUMENT

Table F-12. Example diarylevels input file. F-11
Table F-13. Current microenvironments used by SHEDS. F-11
Table G-1. Inputs to exposure for one-day calculation. G-2
Table G-2. Verification equations for SHEDS Q/A. G-9
Table G-3. Output variables from the exposure calculations. G-12

US EPA ARCHIVE DOCUMENT

ACRONYMS AND ABBREVIATIONS

ADD – average daily absorbed dose

CCA –Chromated Copper Arsenate

CDFs – cumulative distribution functions

CHAD – Consolidated Human Activity Database

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 equivalents

NERL – National Exposure Research Laboratory

NHANES – National Health and Nutrition Examination Survey

OPP – Office of Pesticide Programs

ORD –Office of Research and Development

PBPK – Physically-based pharmacokinetic

PDFs – probability density functions

PK- pharmacokinetic

SA – surface area

SAP – Scientific Advisory Panel

SHEDS – Stochastic Human Exposure and Dose Simulation

UPA – University Partnership Agreement

ug (in SAS printout or variable names) – microgram

yr – year

US EPA ARCHIVE DOCUMENT

1 INTRODUCTION

1.1 Purpose of SHEDS-Multimedia

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), developed by EPA's Office of Research and Development (ORD), National Exposure Research Laboratory (NERL), is 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. On a typical personal computer, a SHEDS variability run of 1000 persons will take about one hour and produces about 500 megabytes of output. These numbers should scale linearly in the number of persons, until limited by computer resources.

Exposure is defined in SHEDS-Multimedia as the contact between a chemical agent and a simulated human target at the skin, lung, and gastrointestinal tract exposure surfaces. Dose is defined in SHEDS-Multimedia as the amount of chemical that enters the target after crossing the exposure surfaces.

SHEDS-Multimedia is a physically-based, probabilistic model that predicts, for user-specified population cohorts, exposures incurred 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 and 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[1] and/or uncertainty in the input parameters.

While the core of SHEDS-Multimedia is the concentration-to-exposure module, there are various options (built-in source-to-concentration module; user-entered time series from other models or field study measurements) for obtaining concentration inputs, and SHEDS-Multimedia exposure outputs can be used as inputs to physically-based pharmacokinetic (PBPK) models.

This Technical Manual describes the algorithms, methodologies, and input and output capabilities of SHEDS-Multimedia version 3, an aggregate residential (non-dietary) model focused on single chemical exposures from inhalation, dermal contact, and non-dietary ingestion. The Technical Manual also includes some detail of the SHEDS SAS code itself (Glen, 2007). The SHEDS-Multimedia version 3 User Guide (Stallings et al., 2007) describes the graphical user interface and the usability aspects of the model. Reviewed collectively, the annotated SHEDS SAS code, Technical Manual, graphical user interface, and User Guide should provide a thorough perspective and understanding of the SHEDS technical and usability aspects. Note that this Technical Manual does not describe a specific application

---

[1] In this manual, "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. Instead of "variability distribution," 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.

of SHEDS-Multimedia version 3 and therefore does not discuss specific model input values. Appendix A describes SHEDS terminology, and Appendix B describes SAS variable nomenclature. The SAS code is organized into modules as described in Appendix C.

SHEDS-Multimedia version 3 does not include the SHEDS-dietary module (Xue, 2007) that estimates exposure from eating food or drinking water; SHEDS-Multimedia version 4 will combine the residential and dietary components in the same model. Plans for SHEDS-Multimedia version 4 also include extending version 3 to include cumulative algorithms (multiple chemicals and their co-occurrence in space and time); a residential fugacity-based source-to-concentration module; enhanced longitudinal activity methods; and a new sensitivity analysis option (Glen et al., 2007). SHEDS-Multimedia version 3 can help answer many useful questions related to populations' aggregate residential exposures for different multimedia chemicals and what factors and pathways are most important. This model can be linked with other tools (e.g., dose estimation models, measurements) for reducing uncertainty in risk assessments. Anticipated primary end users of SHEDS-Multimedia 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), air toxics (SHEDS-ATOX), and wood preservatives (SHEDS-Wood); this manual, however, focuses only on the SHEDS-Multimedia model.

For background information, the SHEDS-Multimedia model development history, general overview of the SHEDS technical methodology, and summary of features and advantages of SHEDS-Multimedia version 3 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, references, and appendices 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 (Zartarian et al., 2000; Hore et al., 2005; Xue et al., 2006; Zartarian et al., 2006) and a number of 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-Multimedia version 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 ORD and OPP. 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-Multimedia (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-Multimedia version 2 to version 3 began. In 2005, the focus of the SHEDS developers returned SHEDS-Multimedia version 3 (the current aggregate version of SHEDS). SHEDS-Multimedia version 3 (SAS) expanded SHEDS-version 2 to include features described in the next section. In 2005 and 2006 NERL and its contractor Alion developed a SHEDS-Multimedia version 3 GUI and the User Guide and Technical Manual. In 2006, SHEDS-Pesticides was renamed to SHEDS-Multimedia since it is designed to accommodate other multimedia, multipathway chemicals. 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 materials have been further tested and prepared for a scheduled review by OPP's FIFRA SAP in August 2007.

SHEDS-Multimedia version 4 and the associated Technical Manual will include algorithms for calculating cumulative exposures and the dietary module, as well as refinements to methods for simulating fate and transport of chemicals in a residential environment (fugacity-based source module) and sensitivity analyses (e.g., Sobol's technique) to identify critical model inputs. The dietary module, and methodologies to extend version 3 to version 4, were reviewed by OPP's August 2007 FIFRA SAP.

1.3 SHEDS-Multimedia Version 3 Overview

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

The primary function of the SHEDS-Multimedia version 3 model (hereafter "SHEDS") is to estimate the exposure of a population to a particular chemical after it has been inhaled, ingested (by hand or object mouthing), or dermally contacted in a residential setting. SHEDS uses the Monte Carlo statistical method to simulate a population of individuals. 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.

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, it is not advisable to pair SHEDS exposure estimates with personal exposure measurements at the level of specific individuals. On a typical personal computer, a SHEDS variability run of 1000 persons will take about one hour and produces about 500 megabytes of output. These numbers should scale linearly in the number of persons, until limited by computer resources.

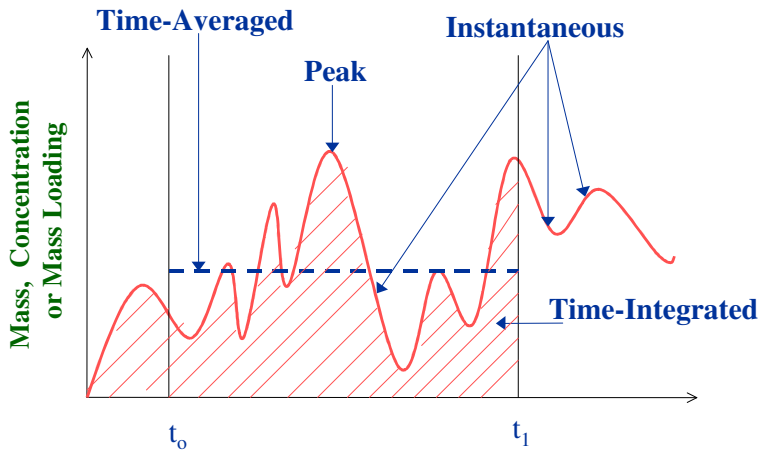


Figure 1-1. Hypothetical exposure profile for an individual.

US EPA ARCHIVE DOCUMENT

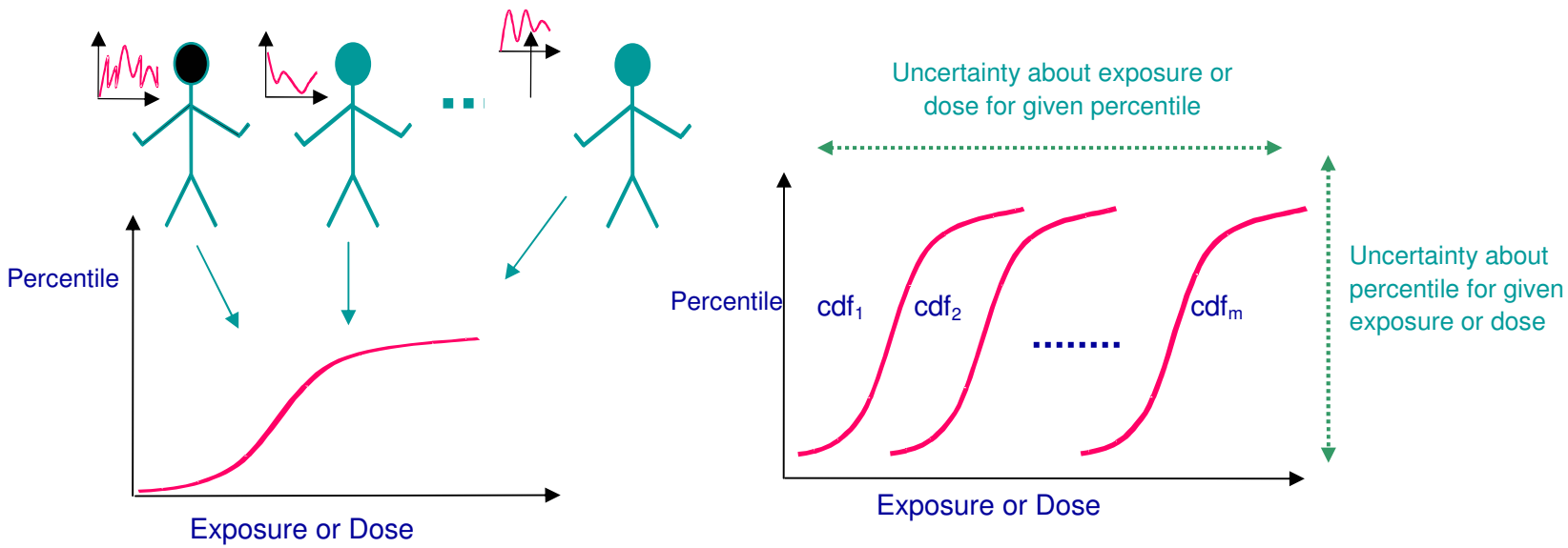
US EPA ARCHIVE DOCUMENT

Variability

- temporal, spatial, or inter-individual differences in the value of an input

Uncertainty

- Lack of knowledge about
  - true value of a fixed but unknown quantity
  - true population distribution for variability



1-stage Monte Carlo

2-stage Monte Carlo

Figure 1-2. Monte Carlo sampling in SHEDS.

In a two-stage run (an "uncertainty run"), 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 of various input variables 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 may take several days to complete.

SHEDS-Multimedia version 3 uses the clock to set random number seeds and does not have the option of reproducing a prior run exactly; this is planned for SHEDS-Multimedia version 4. The ability to reproduce random number streams is essential to the operation of Sobol's method of sensitivity analysis, which is also planned for version 4. Reproducible random number streams will be available even when sensitivity analysis is not being used. If two or more runs are made with the same input distributions, run settings, and random number streams, then the results will be identical.

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. One individual differs from another only because there is stochastic variation when sampling from distributions. For each individual in a SHEDS 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 (using 8 CHAD diaries in SHEDS version 3), which indicates the sequence and duration of activities and locations for that person.
3. Generate concentration time series for each potential contact medium (e.g., indoor air, indoor smooth surfaces, indoor textured surfaces, indoor dust, outdoor air, outdoor lawn, outdoor vegetable garden, outdoor soil, 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 user-specified contact probabilities.
5. Calculate pathway-specific exposure time series for the individual (with exposure equations given starting on page 26), using the results of the prior two steps and user-specified distributions for exposure factors.
6. Generate an approximation for the dose time series, if desired, using the simple built-in pharmacokinetic model in SHEDS (see section starting on page 36).
7. Export exposure time series for use in a PBPK model, or extract desired metrics or summary statistics from the exposure or dose time series.

US EPA ARCHIVE DOCUMENT

To obtain population estimates SHEDS-Multimedia 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 5 possible categories of variables related to exposure and dose (new exposure, running exposure, absorption, dose, and elimination), each divided into 7 pathways (residues to surfaces on hands, residues to surfaces on body, residues to surfaces in GI tract, dust or soil to hands, dust or soil to body, dust or soil to GI tract, air to lungs). Thus, there are 35 basic output variables, not counting sums across pathways. Each of these is reported on each diary event, with typically 12,000-20,000 diary events per year, per individual. When output from just one individual is examined, it is usually plotted as a time series (for any user-specified number of days up to the simulation length). 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 methodology is given in Figure 1-3.

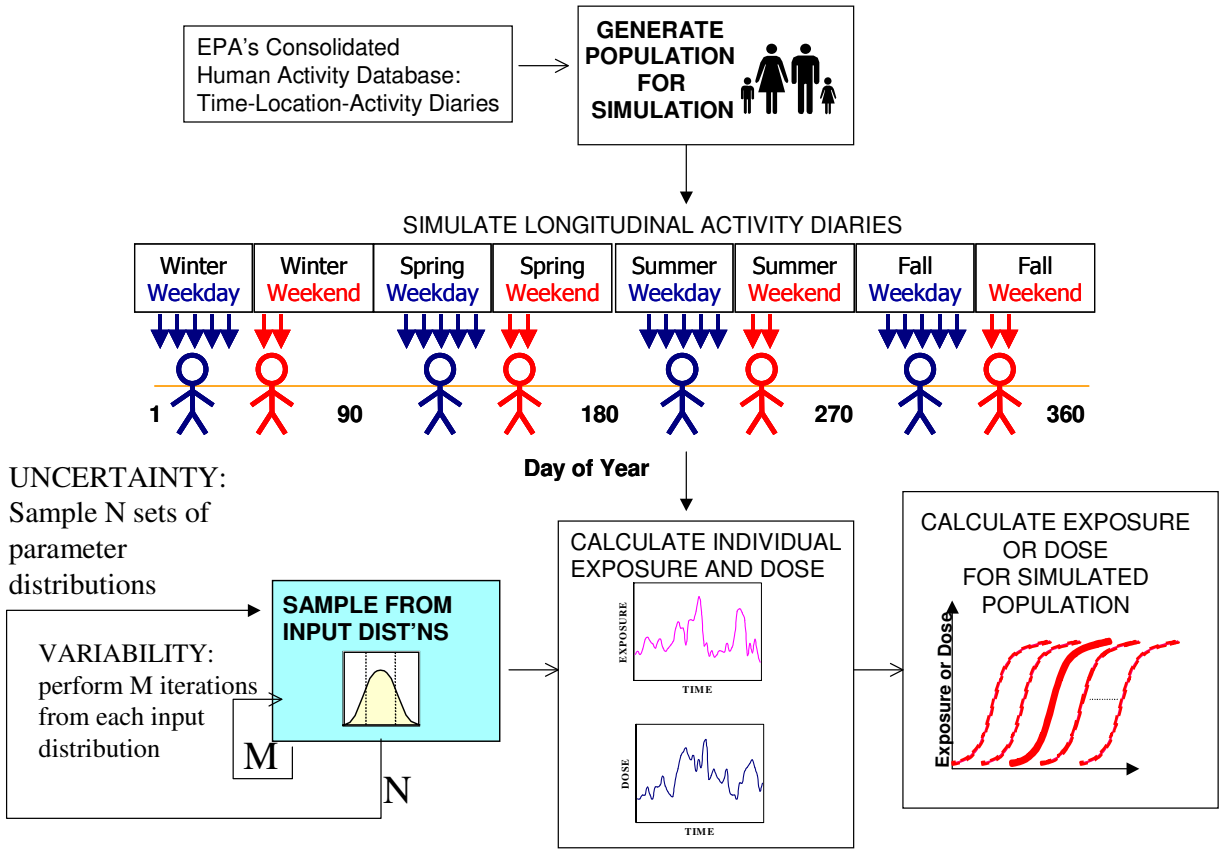


Figure 1-3. Overview of SHEDS methodology.

1.4 Features and Advantages of SHEDS-Multimedia Version 3

Key features of the aggregate version of SHEDS-Multimedia (version 3) 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 (dietary, hand-to-mouth, object-to-mouth, dermal, inhalation) and aggregated (single chemical) exposures/doses;

- 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 for single chemical, multiple application methods;
- exposure scenarios for chemical applications include post-application residential indoor crack and crevice, pets, indoor fogger, lawn, and garden (for different application methods);
- 3 options for source-to-concentration module: read time series; sample from distributions based on time since application; use built-in decay/dispersion model;
- 2 options for exposure-to-dose: export exposure time series to PBPK model; built-in simple PK model;
- 2 options for soil/dust ingestion: direct ingestion; hand-to-soil/dust-to-mouth;
- 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; concentrations/residues for surfaces (smooth or textured), air, dust, as well as outdoor soil, lawn, and vegetable garden;
- separates dermal hand and body exposure profiles and links hand-to-mouth ingestion with dermal exposure; and
- accounts for sequential activities, dermal replenishment, and removal processes (hand-to-mouth, bathing, hand washing, absorption).

US EPA ARCHIVE DOCUMENT

2 MODEL STRUCTURE AND APPROACH

This section covers the SHEDS 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. First simulation of individuals will be described below, and then population simulation.

SHEDS-Multimedia version 3 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. 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-Multimedia 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 types 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 expected to be utilized in version 3 and 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 categories; basal ventilation rate equation; body surface and hand surface area equations; maximum METS, maximum oxygen debt, and recovery time formulae; the re-entry time bins; 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 User Guide for explicit detail on using the SHEDS 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.

If the concentration time series option is not being used, then 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 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 (due to lack of available exposure-relevant information on aging an individual). If the concentration time series option is used, then the simulation period is defined by the 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 run of 100,000 persons impractical, if this number were desired one could perform 10 runs of 10,000 persons each. However, if this is desired, such analysis must be performed in SAS, 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 three possible 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 two options first require the determination of chemical usage. 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 post-application 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.

Unless chemical concentration time series are entered, the model must assign chemical usage characteristics to each individual. The user selects one or more "application types," which are combinations of chemical formulation, method of application, and target location. These are described in more detail below. Finally, there are two methods for determining the dates of chemical usage, called 'UserDates' and ModelDates.' Either of these may be used in combination with either the decay/dispersion method or the post-application method. In the 'UserDates' option, the user specifies which days of the simulation period have chemical usage, for each application type. All persons in the model run then share the same set of usage dates. The other option is 'ModelDates,' in which the model randomly selects chemical usage dates for each application type, based on user-defined probabilities. In 'ModelDates,' each individual will have different usage patterns. It may be possible for some individuals to have no chemical usage at all.

The user next decides whether soil ingestion is to be modeled directly (i.e., using user-specified soil ingestion rates) or via the mechanism of hand-to-mouth transfer. The maximum duration for a diary event is also specified by the user. 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 concentration time series (which may depend on chemical 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 with Monte Carlo sampling to simulate a population from which output information can be obtained.

2.2 Creating a Simulated Individual

2.2.1 Age Groups

SHEDS uses age groups for three purposes. First, the model samples activity diaries from pools spanning the age groups. 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. Third, the results from a model run may be grouped by age.

For the first two purposes, SHEDS currently uses the following age groupings: 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.

Because of lack of available information, the SHEDS version 3 residential model does not specifically model infants less than one year of age. 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 Child-Specific Exposure Factors Handbook. For the third purpose of viewing SHEDS population exposure results, each simulated individual has a specific age and gender, and the SHEDS outputs can be customized into user preferred categories (e.g., 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).

2.2.2 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). The user may select any or all of the 20 age-gender cohorts to be in the target population. SHEDS randomly assigns a particular age and gender to each individual, and these age and gender 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 all persons of a given age will also show age dependence.

2.2.3 Assigning other person-level variables

The 'Persons' module in the SHEDS code assigns age and gender to individuals, along with other "person-level" input variables (those 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. 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. At present the weights are not capped at minimum or maximum values. However, a sample 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. 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. Regression equations were developed for body weight, height, and surface area that preserve correlations among those parameters for a given person.

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 probability (a point distribution). A yes/no test is performed for each of these questions, for each individual. The defaults are appropriate for the population at large. If the target population is intended to be pet owners only, then one would set has_pet_p=1, to ensure that all the randomly generated persons have pets. Similar remarks apply to the lawn and garden.

By editing the "variables" input data set in the SAS code, the user may increase or decrease the number of variables sampled once per person, and thereby alter the variables appearing on the "persons" data set. This feature is not supported by the SHEDS interface, as it could lead to problems with undefined quantities in the code.

The "Persons" module in SHEDS is unique in that it assigns data to all the individuals in the model run at one time, before calculating detailed activities and exposures for any one individual. This is done primarily for reasons of efficiency. The "persons" output file has 46 variables when following the default settings, and can store information on more than 2,500 persons per megabyte. In practice, SHEDS runs will not contain enough individuals to cause problems storing all this information in one data set.

2.3 Generating Longitudinal Activity Patterns for a Simulated Individual

2.3.1 Selecting Activity Diaries

The 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). The diaries are matched to the SHEDS individual by age-gender cohort, and by season and day-type (i.e., weekend or weekday).

There are four studies in CHAD that provide children's diaries: University of Michigan study; (2) National Human Activity Pattern Survey; California children's study; and Cincinnati 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 adult's study, and most of the diaries in the Cincinnati study and 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 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 (not counting those less than 1 year), 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 CHAD diaries cover just one or two days, a method is needed for combining several CHAD diaries to cover a longer period. For each individual, SHEDS-Multimedia version 3 constructs a year-long diary by sampling 8 CHAD diaries, one weekend and one weekday diary from each of 4 seasons (Figure 2-1). 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. It should be noted that while the CHAD diaries provide information on general location (such as inside one's house) and activity (such as eating dinner), they do not provide enough detail to determine the precise extent of the contact with chemical-containing media. This level of detail is provided by randomly sampling the distributions for the relevant exposure factors, as specified by the user. So while the same CHAD diaries are repeated many times for one person, the exposure events are unique on each simulation day.

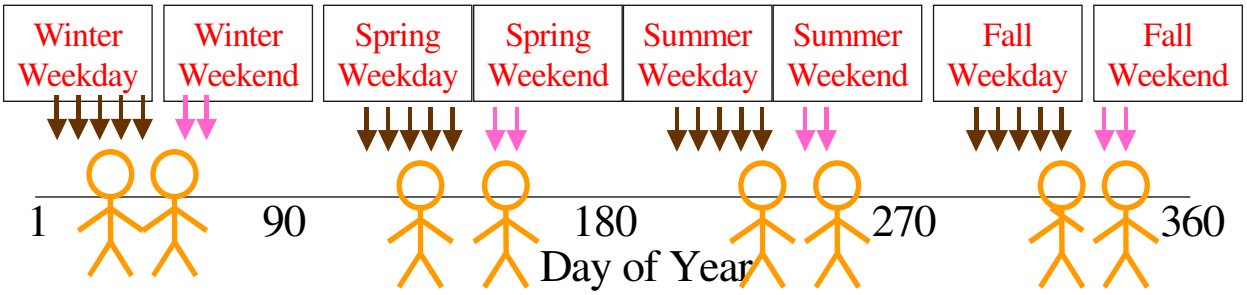


Figure 2-1. SHEDS approach for simulating one-year activity patterns.

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 contacts with treated media are the same from day to day. As explained below (see "Determination of Contact Events" section), a random subset of the potential contact time becomes simulated contact time, and this subset varies from day to day.

2.4 Simulating an Individual's Contact with Scenario-Relevant Media

2.4.1 Contact Media

In SHEDS, humans (targets) are exposed to a chemical (agent) by coming into contact with various contact media. There are three basic types of media considered in SHEDS that are referred to as "surface," "air," and "matter." A given model run may have several examples of each type.

"Surface" media in SHEDS (e.g., smooth surfaces such as vinyl floors; textured surfaces such as carpeted floors) represent their loadings in units of micrograms of the target chemical per square centimeter of surface. This residue loading can be entered by the user as total or dislodgeable surface residues (the user must be careful to enter the appropriate corresponding residue transfer factor). This surface loading may diminish naturally 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 SHEDS the exposure surface for dermal contact with surface media is the human skin surface.

"Air" media have concentrations measured in units of micrograms of chemical per cubic meter of air ($\mu g/m^3$). The amount of inhalation exposure depends on the air concentrations and the simulated person's ventilation rate (units of cubic meters of air per hour), which depends on the nature of the activity and the metabolic rate of the person. In SHEDS the exposure surface for inhalation exposure to air media is the lung lining (dose to the blood occurs after the agent crosses the lung lining).

"Matter" media in SHEDS refer to soil or dust, with concentrations measured in units of micrograms of chemical per gram of matter ($\mu g/g$). The distinction between "soil" and "dust" in SHEDS is one of source location: if it is encountered outdoors it is soil; if encountered indoors it is dust. In SHEDS, the dermal contact with soil or dust is measured by an "adherence rate," expressed in units of milligrams of matter per square centimeter of exposed skin, per hour. Although not currently supported by the GUI, by directly editing the 'distributions' input file one can have the model use different adherence rates for soil and for dust. SHEDS may also allow direct ingestion of dust and soil, as discussed below. In SHEDS the exposure surface for ingestion exposure to matter media is the lining of

the gastrointestinal tract (dose to the blood occurs after the agent crosses the GI tract lining). While food and water are other types of "matter" media that use similar units for concentration and might contribute to exposure, these are not integrated into the SHEDS-Multimedia version 3 model.

The list of contact media is data-driven; the model processes all the media specified in the input files. In SHEDS version 3 the default list is the following:

Table 2-1. Default list of contact media in SHEDS-multimedia version 3.

Variable	Type	Description
Tair	air	air in treated room
Tdust	matter	dust in treated room
Thard	surface	smooth (hard) surface residues in treated room
Tsoft	surface	textured (soft) surface residues in treated room
Uair	air	air in untreated room
Udust	matter	dust in untreated room
Uhard	surface	smooth surface residues in untreated room
Usoft	surface	textured surface residues in untreated room
Yair	air	air outside residence
Ysoil	matter	soil in treated outdoor locations
Ylawn	surface	surface residues on lawn
Yveg	surface	surface residues in vegetable garden
Pet	surface	residues on pet fur

2.4.2 Determination of contact events

The longitudinal diary assembled from the one-day CHAD diaries gives the general activity, location, and duration information that characterizes the time-activity behavior of the simulated individual. However, the SHEDS model requires knowledge of details such as whether or not particular surfaces are touched. This level of detail is not recorded in CHAD and must be stochastically generated and added to the diaries. These operations in SHEDS are handled by the SAS code 'DiaryDetailing' module, which performs two main tasks: (1) determining the extra bathing (since CHAD underreports bathing) and the hand washing events; and (2) generating the simulated contacts with surfaces, air, and matter media.

In addition, this module subdivides any diary events that exceed the maximum permitted duration (as input by the user), and fills in any missing activities or locations on the CHAD diaries themselves. The maximum diary event duration in CHAD is 60 minutes, but in SHEDS the user may optionally request a shorter time (based on experience with running SHEDS, shorter times have minimal effect on exposure). Diary events over the maximum are divided into pieces with the same activity and location, but possibly with different contacts. Missing activities and locations on the CHAD diaries are simply set equal to the prior non-missing values. The default diary file has the following criteria imposed on it: <1 hour of missing locations and <2 hours of missing activities; these can be adjusted through the SAS input

files. The determination of extra bathing and hand washing events is described in the section titled "Activity-related exposure factors" on page 58.

For each diary event, SHEDS first determines the appropriate microenvironment ("micro"), then the sub-microenvironment ("sub-micro"), then the contacts with various media. The SAS code variable "micro" reflects the microenvironment where the individual is located. There are five micros in SHEDS:

Table 2-2. Micro environments (micros) in SHEDS.

Variable Name	Micro
in_home	inside a residence
in_veh	in a vehicle
in_oth	inside another building
out_home	outside around home
out_oth	outside away from home

The mapping from CHAD codes to SHEDS micros is in the SAS input file 'diarylocs.' This file can be edited in SAS, but not directly via the SHEDS interface. The mapping from CHAD codes to SHEDS micros is deterministic (no stochastic variation). By contrast, the sub-micro is selected randomly. Inside the residence, the sub-micro may be 'treated room,' 'untreated room,' or 'far away.' The input variable 'probT_in_home' gives the probability that an in_home event is in a treated room. The variable 'probU_in_home' gives the probability for being in an untreated room. Finally, 'probN_in_home' gives the probability for being in neither (that is, too far from a treated area for any chemical to be present). Note that the meaning of "treated" is independent of time in this context; it could mean being treated in the future.

Once the sub-micro or location has been selected for the given diary event, SHEDS chooses the media that are contacted. SHEDS permits only one chemical-containing medium (at most) of each type (air, matter, and surface) to be contacted on a single diary event. In each sub-micro, there are up to four possible outcomes for surface contact, two for matter, and just one for air. Air must always be breathed, although some locations may have zero chemical concentration in the air. For matter, the decision is a yes/no test based on user-supplied probabilities. Indoors, the person may contact dust, or no matter at all on that diary event. Outdoors, the person may contact soil, or no matter at all. As mentioned above, in SHEDS the terms "soil" and "dust" refer to their location, not their physical or chemical properties. "Soil" is matter contacted outdoors, while "dust" is matter contacted indoors.

Suppose a given diary event randomly is assigned to be in a treated area. Following Table 2-1 the surface could be 'Thard,' 'Tsoft,' 'Pet,' or 'none' (see the previous section on contact media for the definitions of these variables). The outcome is decided randomly on each diary event, using the standard sampling methodology for probability vectors. Every diary event is given a set of contacted media comprising one surface medium, one matter medium, and one air medium, with "no contact" being an option in the first two cases.

US EPA ARCHIVE DOCUMENT

2.5 Generating Concentration Time Series for the Contact Media

2.5.1 Chemical usage patterns (applications)

Unless the user selects the 'TimeSeries' input option for concentrations, the environmental chemical concentrations will be the result of chemical usage patterns. There are two methods of specifying chemical usage patterns. One is for the user to specify the calendar dates when chemicals are applied (the 'UserDates' option). Here, all simulated individuals will share the same usage patterns. However, the fraction of the house treated and the mass of chemical applied may vary from person to person. The UserDates option is useful, for example, for estimating the variability in exposures that may follow an application. All persons in the model run may experience an application on day 1 (or on day 0), and the comparison of individuals on similar days of their time series provides the exposure distribution for that point in time. The UserDates option also supports multiple applications of one type, as well as multiple types. For the UserDates 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.

The second method for determining chemical usage patterns is called the 'ModelDates' option. Here, SHEDS randomly generates chemical usage frequencies and dates according to user-specified rules. SHEDS does not contain any proprietary information, but users may supply such information as input data at their discretion. Output data analyses from ModelDates runs often differ from those of UserDates runs, since individuals are not necessarily exposed at the same times. For ModelDates, summary statistics over each time series are computed first, then the results are compared across the population. For example, one may be interested in the mean daily exposure, maximum single daily exposure, or total number of days over some exposure threshold. Such quantities are not directly dependent on the exact timing of the chemical usage.

Apart from the determination of chemical usage frequencies and application dates, the UserDates and ModelDates options are very similar. Most sections discussed below apply to both methods. Exceptions are noted, with an explanation of the differences.

2.5.1.1 Application types

An "application type" refers to a specific combination of chemical formulation, application method, and target location. An example of target location is the lawn. Pesticides may be applied to the lawn either in granular form (e.g., using a mechanical spreader), or liquid form (e.g., using a pressure sprayer). Other alternatives may be possible as well. Indoors, there are even more choices for application types. The user selects the application types for a particular model run from a master list provided by the interface. In batch runs, it is possible to add new application types without changing the SHEDS code, but this is fairly difficult and is not recommended for most users, nor is it supported by the interface.

The application types currently supported in the interface in SHEDS-Multimedia version 3 are:

Table 2-3. Chemical application types in SHEDS.

Type	Description
InCCaer	Indoor crack and crevice (aerosol)
InCCliq	Indoor crack and crevice (liquid)
InCarRm	Indoor carpet and room (aerosol)
InFIK	Indoor flying insect killer (aerosol)
InFog	Indoor fogger (broadcast)
LawnGra	Lawn (granular - push spreader)
LawnLiq	Lawn (liquid - handwand)
PetLiq	Pet treatment (liquid)
PetSpot	Pet treatment (spot-on)
VegDust	Vegetable garden (dust, powder)

The user specifies which application types are being considered in each model run. In a UserDates run, all of the selected types will be "active" for each person; that is, they will definitely occur for every individual. In a ModelDates run, the model randomly determines which of the selected application types are active for each person. It is possible for very few (if any) persons to experience the entire selected set.

The relationship between application types and contact media is complicated. Each application type may affect several media, and each contact medium may be affected by several application types. For example, an indoor fogger may affect Tair, Tdust, Thard, and Tsoft directly, and affect Uair, Udust, Uhard, and Usoft indirectly. Conversely, Tair may be affected by any of the indoor application types (InCCaer, InCCliq, InCarRm, InFIK, or InFog on the default list). In addition, all contact media may have background concentrations that occur even in the absence of an application type. More details on the relationship between application types and contact media are presented in the section on "Media concentration generation methods."

2.5.1.2 Active applications and application frequency

In the UserDates option, all of the application types selected for the model run are active for all individuals. Furthermore, they occur with the same frequency and on the same dates for all individuals. The ModelDates option requires significantly more information to determine the application activity, frequency, and dates. A loop over application types makes the following decisions for each person:

Is the application type active for this person?

If active, how often does this application type occur?

On which specific dates does this application type occur?

The following briefly describes how the code answers these questions. First, not all application types can occur or will occur for each person. The two lawn application types can only occur for persons for whom the variable Has_lawn is set to 1 (which represents "yes"). This is determined on the Persons

data set, as part of the demographic module, using the input variable Has_lawn_p. The Has_lawn_p variable is a probability, usually set to a point value such as 0.75. Then each individual would have a 75% chance of having Has_lawn=1, and a 25% chance for Has_lawn=0. Note that the variable Has_lawn varies from person to person, whereas Has_lawn_p does not. Similar rules apply for the two pet application types (which require Has_pet=1) and for the Vegetable Garden application (which requires Has_Garden=1).

Even when a person has the specified requirement or prerequisite, s/he will not always use the chemical on the location in question. Only a fraction of persons with lawns will use the target chemical on them in a given year. The SAS variable that controls this is called "user_pr", which is specified for each application type. It represents the likelihood of actively using the application type, given that the appropriate requirement has been met (having a lawn, garden, or pet, as the case may be). Thus, in the above example, if user_pr = 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). Even in the ModelDates option, one can make treatment a certainty by setting Has_lawn_p=1 and user_pr=1.

Once an application is found to be active for a given person, the frequency of application is determined. As discussed earlier, the application frequency is always determined on an annual basis regardless of the simulation length, and the application calendar must be extended at least 30 days prior to the start of the simulation. In many cases, the application calendar extends for two full years.

For each application type there is a 'NumApps' vector which gives the probabilities for 1, 2, 3, …N applications per year (N=12 using interface and N=52 for batch mode). One draw is made from the appropriate vector for each active application type (for each person). Suppose the result is 2, and the application calendar happens to contain two years. This implies that the given treatment type is used four times during that two year period. Then, 4 specific application dates must be chosen for this application type, using the rules in the next section. Note that these 4 applications could be assigned to dates before, during, or even after the actual simulation period. A separate draw of application frequency is made for each active application type.

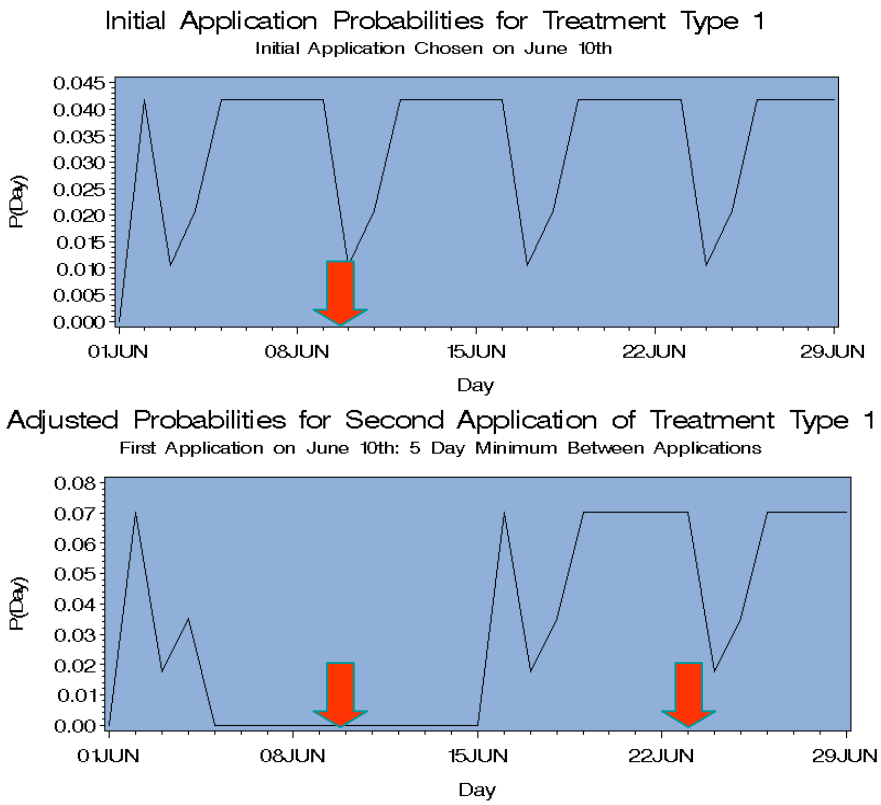
2.5.1.3 Application dates and co-occurrence

Once the frequency of each application type 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 application types are made (Figure 2-2 and Figure 2-3). These adjustments are made based on the values of the influence factors and influence widths, discussed below.

US EPA ARCHIVE DOCUMENT

For a Single Treatment Type the User Specifies

- Treatment type and priority for co-occurrence
- Probabilities of applications based on month and day of week



Initial probabilities for treatment type 1. First Application date chosen on June 10th.

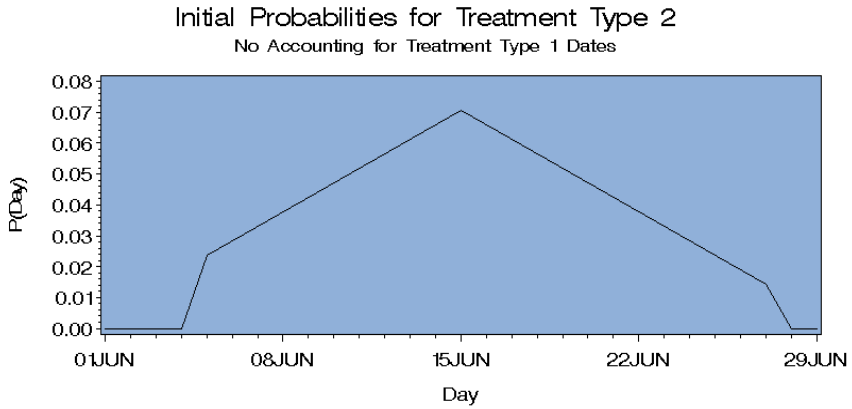
Probabilities adjusted for initial June 10th application date assuming a 5 day minimum time between applications. Second date chosen on June 23rd.

Figure 2-2. Choosing application dates for a single treatment type.

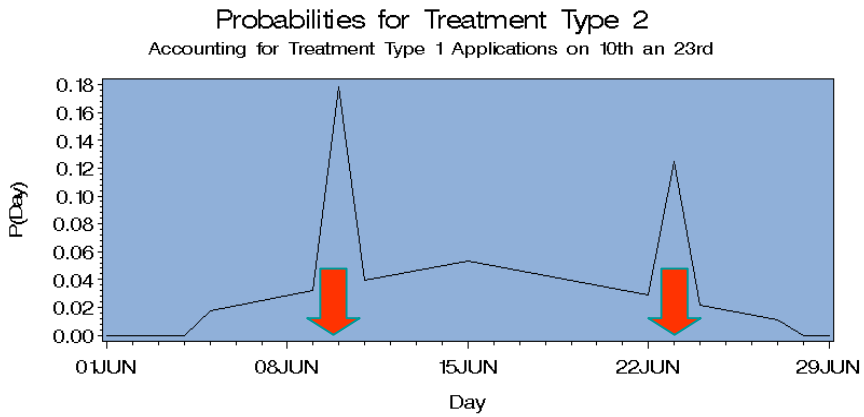
US EPA ARCHIVE DOCUMENT

User Specifies Treatment Types and Relationships

- All quantities needed to define individual treatment types
- Priority among treatment types (type 1 was first priority, type 2 is second, …)
- Influence factor and influence width for each pair of related treatment types



Initial probabilities for treatment type 2 not accounting for treatment type 1 application dates.



Probabilities adjusted for treatment type 1 application dates, June 10th and 23rd. These probabilities will be used to pick first date for treatment type 2.

Figure 2-3. Co-occurrence of treatment types, choosing application dates for treatment type 2.

The adjustments are calculated on a single pass over the application types rather than iteratively, so the adjustments take into consideration only the dates that have already been assigned to this person. Thus, the first application type on the list never has adjustments for other application types. The second application type might (or might not) have adjustments that depend on the first. The third can be adjusted for either of the first two, and so on. To accomplish this, the user assigns a priority order to the application types, that indicates which are processed first. The application date probability adjustment for other application types can only be specified for application types that occur further down the priority list than the ones on which they depend.

The application probability adjustment takes the form of a probability multiplier (influence factor) that is applied to all days within the specified number of days (influence width) of a treatment of the indicated type. For example, suppose application type #1 occurs on June 10, application type #2 occurs on June 20, and application type #3 occurs on June 30. The user-specified co-occurrence rules are set so that application type #4 is only half (i. e., influence factor=0.5) as likely to occur if either types #1 or #2 (or both) occur within an influence width of 7 days. This means that any date less than 7 days from a #1 or #2 application date has its probability for application type #4 cut in half. Specifically, this means all dates from June 4 through June 26 inclusive have their probabilities halved. The dates around June 30 are unaffected because application type #3 was not included in the list that affects type #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. Then one date is selected at random from this large probability vector. This becomes a definite date for an application of type #4. Suppose the frequency of applications of type #4 was determined to be four. Then the remaining three dates still need to be chosen, but first the probabilities must be adjusted again for blackout periods. Suppose the first date for #4 was found to be September 15 and the blackout period for application type #4 is 14 days. This means that no two applications of type #4 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 applications of that type have been assigned, or all possible dates have been blacked out. In the latter case, no more applications of that type are assigned, so the number realized is less than the number specified by the NumApps value.

Each application is given a specific hour of the day as well as a date. The hour is selected from either the indoor TimeOfDay vector (indoor or pet applications) or the outdoor TimeOf Day vector (lawn or garden applications). Each application is given its own hour. Two applications of different type may occur on the same calendar day, and may (accidentally) occur on the same hour, although this generally would be extremely rare.

Application dates that fall before the simulation period may nonetheless produce concentrations that persist into the simulation period, so they are potentially important. Application dates beyond the end of the simulation period have no effect on exposure, but these “use up” some of the available number of applications. This method allows the user to obtain and set NumApps and Month data that do not need to be changed simply because the simulation period is changed. If the application period matched the simulation period exactly, then the NumApps vector and the Month vector would have to refer to that period, at least implicitly. For example, suppose the user is running a simulation for the month of March alone. What should the NumApps distribution be for just March, given the annual distribution? Instead of attempting to answer this, SHEDS uses the simpler approach of assigning applications to a full year

US EPA ARCHIVE DOCUMENT

using the full NumApps. This also has the advantage that the March exposure can take into account residual chemical from applications in January or February.

In the UserDates option, one or more dates are specified for each application type in a comma-separated list. For example, the list (1, 91, 181, 271) indicates that an application occurs on the first day of the simulation period (which is always numbered as "day 1"), and also on days 91, 181, and 271. It is possible to include zero or negative numbers in this list, to force applications prior to the simulation period. The numbers in each list should be in ascending order. In UserDates, all applications of a given type occur at the same clock hour for all persons, ensuring that the time series are "in phase" in terms of hours since application. If multiple application types are used, each one may have its own clock hour.

2.5.1.4 Re-entry prohibition times

Often, people may choose to avoid (or are told to avoid) entering a treated area shortly after a chemical application is made. In SHEDS, this is called the re-entry time prohibition. Two probability vectors are used, one for indoor applications and the other for outdoors. Each vector consists of nine probabilities. These correspond to the following categories: 0, 1, 2, 3-4, 5-6, 7-12, 13-24, 25-48, 48-96 hours. For example, the vector (0.5,0,0,0,0.5,0,0,0,0) means a 50% chance of no prohibition, and a 50% chance of a 5-6 hour prohibition. The first three categories are exact amounts of time, while the last six represent ranges. If a range is selected then each hour in the range is equally likely, so in this example it means a 25% chance for 5 hours and a 25% chance for 6 hours.

It should be noted that the first contact of the simulated person with the pollutant is governed by the activity diary. Thus, if the lawn is treated at 10 a.m., it cannot be contacted until the first 'outdoor-residence' activity event after 10.a.m. on that diary day (if there are any at all). The re-entry time prohibition is therefore a minimum only. If there were a 6 hour prohibition, but the diary had an outdoor-residence event at 1 p.m., then this would automatically become a non-contact event. If the same diary happened to be in use for the next day as well, then a random check would be made to determine if this 'outdoor-residence' event involved any lawn contact or not at 1 p.m. on the second day.

If the user does not want a re-entry prohibition (e.g., with mosquito repellants), then all the probability should be placed in the first element, as (1,0,0,0,0,0,0,0,0).

2.5.2 Media concentration generation methods

SHEDS version 3 has three options for generating exposure concentration time series for the various contact media: decay/dispersion, post-application distributions, 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 concentrations are determined, the output from the concentration module in SHEDS is similar; one concentration time series for each contact medium.

The concentration time series are determined hourly. For a standard 365-day simulation, this amounts to 8760 hourly values. 'Hour 1' of this time series corresponds to the period from midnight to 1 a.m. on day 1 of the simulation. If a new application occurs at 10 a.m. on day 2, then this is the start of hour 35, so the 35th value of the hourly time series will likely be much larger than the preceding values. The amount of the increase depends on the selected source-to-concentration approach.

2.5.2.1 Decay/Dispersion Method

The decay/dispersion method uses first-order decay logic for concentrations in treated areas and uses dispersion calculations 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 and the decay rates for each of the treated media. The user specifies a decay rate in units of fraction lost per day. Thus, a decay rate of 0.15 means that the concentration drops by 15% of the current level, per day. Since the current level declines over time (in the absence of new applications), the concentration time series flattens out as it approaches zero.

The model converts the daily decay rate D_d to an hourly decay rate D, based on the relationship $(1-D_d) = (1-D)^{24}$, or equivalently, $\log(1-D_d) = 24 \log(1-D)$, which can be solved to give $D = 1 - \mathrm{Exp}[(1/24)\,\mathrm{Log}(1-D_d)]$. If the daily fraction lost is much less than one, then the hourly fraction lost is close to 1/24 of the daily amount. The concentration at hour 't' is given by the prior value times (1-D), unless a new application occurs.

Equation 2-1

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

where C_t = concentration at hour 't'

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

D = decay rate (constant over time series), 0<D<1, units (1/hour)

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

The treated media are indicated by a prefix 'T' in their names. Thus, 'Tair', 'Tdust', 'Thard', and 'Tsoft' are media in the treated room(s). For an indoor application type using the decay logic, there are 4 instances of the above equation (one for each treated contact medium), and 8 input variables (distributions for D and A_t for each contact medium).

Each of the 4 "treated" indoor media has an "untreated" equivalent, with prefix "U" instead of "T". The equation for the concentration in each untreated medium is:

Equation 2-2

$$\mathbf{C_t = C_{t-1}\ (1 - D) + (S_t - C_{t-1})\ F}$$

where
C_t = concentration at hour 't'
C_{t-1} = concentration at hour 't-1'
D = decay rate (constant over time series), 0<D<1, units (1/hour)
S_t = concentration in corresponding medium in treated room at hour 't'
F = fraction of source dispersing into this medium per hour.

D is the decay rate, which is assumed to be equal to the decay rate in the treated medium (which is logical since both the chemical and the physical nature of the medium are the same as for the treated case). The concentrations S_t are determined using the prior equation for treated areas. The only new variable in the above equation is therefore F, the fraction of the source dispersing into the untreated medium per hour. F is liable to be difficult for the user to determine, therefore SHEDS asks the user for the concentration ratio between the untreated and treated media, at the time when the untreated concentration reaches its maximum. Typical values for this variable may be on the order of 10% or 0.10. This ratio should be

US EPA ARCHIVE DOCUMENT

obtainable from field studies that make measurements in both treated and untreated rooms. The fraction F in the above equation can be calculated as the product of the decay rate D and the untreated/treated ratio UTratio. Thus, if the decay rate is 12% per day or about 0.5% per hour and UTratio is 10%, then F is 0.005 x 0.10 = 0.0005 per hour.

The concentration in treated media spikes upward suddenly at the time of application and declines monotonically thereafter until the next application. By contrast, untreated media exhibit no sudden jumps. At the time of an application, the concentration in untreated media starts to rise gradually, eventually reaching a peak, and then declining. In the long run, this decline is at a rate close to the decay rate in the treated medium.

Outdoor applications may also use the decay logic, although there are no untreated versions of outdoor contact media in the default inputs. While untreated media always have relatively low concentrations, the potential for long contact times exists indoors. For example, bedrooms are often untreated, depending on the nature of the chemical and the problem it is being used to address. Outdoors, the contact times are relatively short, so contact with untreated media is generally not considered to be a significant source of exposure.

When several application types affect the same contact medium, a separate concentration time series is calculated for each application type. Under the assumption of independence, all the time series affecting the same medium are added together, including a time series for the background concentration (if any). For example, the final time series for Tair is given by the term-by-term sum of the Tair time series resulting from each application type independently.

2.5.2.2 Post-Application Method

An alternative to the decay/dispersion approach is the post-application 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 application. In some practical measurement situations, this sort of input data might be more readily available than the decay/dispersion inputs.

When an application occurs, a percentile is chosen at random (uniformly). Small percentiles correspond to light usages of the chemical, while larger percentiles represent heavier usages. On the application 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. When the time is right, 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 application, and the concentrations cannot fall as a result of a new application. If either exception occurs, then the prior concentration continues to be used.

The post-application method does not sum the results from multiple applications of the same type. This was deemed inappropriate, since the measured concentrations being used as the basis for input for the four time periods presumably already include the effects of previous applications. On the other hand, if different application types are present and affect the same contact medium, then the concentrations resulting from each are added.

The post-application method does not directly relate the treated and untreated media. Therefore, if the user wishes to have non-zero concentrations in the untreated media, then distributions for these must be specified as well. Therefore, as many as 32 distributions may be required for each application type when using the post-application logic (8 affected media for 4 time periods each).

2.5.2.3 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. The user may supply multiple examples of each contact medium, corresponding to a set of houses. This data set has one record per day. The list of variables must include one called "date" which indicates the date for each record in SAS date format. It also requires variables whose names are in the form "medium_house#". For example, if there were data from three houses, then the air concentrations in the treated room would be under the variables named "ATair_1", "ATair_2", and "ATair_3". The corresponding dust concentrations would be "MTdust_1", "MTdust_2", and "MTdust_3". The first letter of these names indicates the type of media: A for air, S for surfaces, or M for matter. The types indicate the necessary units: ($\mu g/m^3$) for air, ($\mu g/cm^2$) for surface residues, or ($\mu g/g$) for matter. The units do not explicitly appear on the input files; the user must ensure that they are correct. The suffix 1, 2, or 3 indicates the house number. SHEDS selects one house at random for each person, and uses all the concentrations from that house together. Other variables may be supplied 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 must have no gaps (days with missing data) between their start and stop dates. All of the time series for any given house must have the same start and stop dates. As an example, suppose there are just three houses and two media. The following input file would be valid:

Table 2-4. Example input file for concentration time series.

Date	Thard_1	Tsoft_1	Thard_2	Tsoft_2	Thard_3	Tsoft_3
14JUL2006	.	.	5.4	1.2	.	.
15JUL2006	.	.	3.3	0.6	.	.
16JUL2006	6.8	4.2	1.7	0.4	.	.
17JUL2006	5.8	4.7	.	.	.	.
18JUL2006	3.7	2.6	.	.	.	.
21JUL2006	.	.	.	.	7.4	5.2
22JUL2006	.	.	.	.	6.1	3.5

Note that there are two days missing between the end of house 1 and the start of house 3. These records may either be present or omitted on the input file; it does not matter. If present, then of course all the concentrations on these days would be indicated by periods. 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-Multimedia job, the user specifies a start date and a number of days, defining the simulation period. For the UserDates and ModelDates 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 17, 2006 and the number of days to 7. Then only the last four records in Table 2-4 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 each simulation would be two days long, but in general the number of days may differ between houses.

2.6 Generating Exposure Time Series for Individuals

2.6.1 SHEDS equations for exposure by pathway

The estimation of exposure is the central function of the SHEDS model. Exposure is defined in SHEDS-Multimedia as the contact between a chemical agent and a human target at the skin, lung, and gastrointestinal (GI) tract exposure surfaces. The skin exposure surface is separated in SHEDS into hands and the rest of the body. The time series of exposure preserve within-day peaks and variability as a person moves through time (Figure 2-4 and Figure 2-5).

These exposure profiles can yield toxicologically relevant dose profiles (especially for chemicals with short half-life), and ultimately, improved risk estimates. Exposure is calculated for each diary event, resulting in pathway-specific time series for each individual. The skin (hands and skin on rest of body) and GI tract may be exposed to either residues (that is, chemical which is transferred from surface residues) or to matter (dust or soil). The lungs are exposed only to air. Thus, SHEDS tracks seven pathways for exposure:

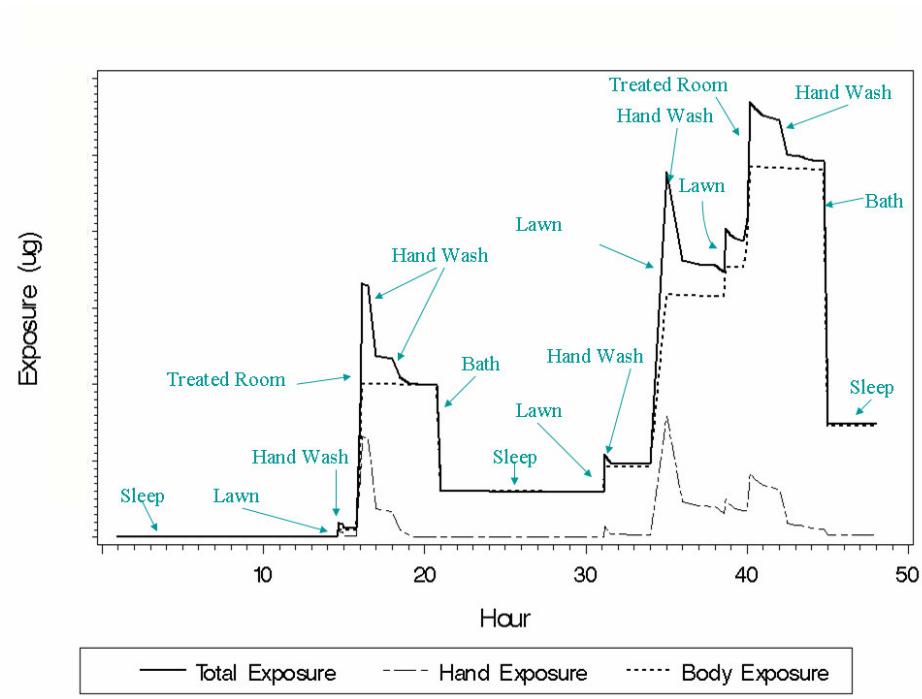


Figure 2-4. Example time profile of dermal exposure from surface residues.

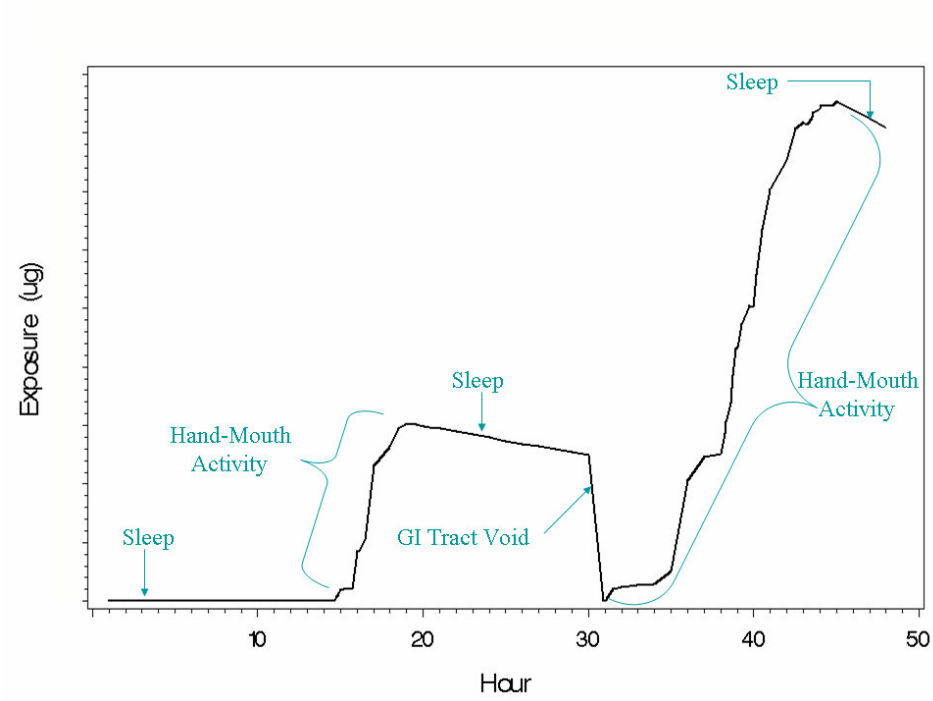


Figure 2-5. Example time profile of GI tract exposure from surface residues.

In all cases, SHEDS internally tracks the mass of the chemical agent present, in micrograms. On the daily output files, these variables may be expressed either in micrograms per day (µg/day), or milligrams per kilogram body weight per day (mg/kg/day).

Table 2-5. Exposure pathways in SHEDS.

Symbol	Description
HR	residues from surfaces on hands
BR	residues from surfaces on body
GR	residues from surfaces in GI tract
HM	dust or soil on hands
BM	dust or soil on body
GM	dust or soil in GI tract
LA	air in lungs

SHEDS retains two measures of exposure. 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 nonzero 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

US EPA ARCHIVE DOCUMENT

exposure for a given diary; it merely reduces the running exposure. The seven variables for new exposure in the SHEDS SAS code are expHR, expBR, expGR, expHM, expBM, expGM, expLA. The seven running exposure variables are similar except that "exp" is replaced by "ldg" in each case.

The distinction between new exposure and running exposure is important. New exposures may be summed across diary events; the total new exposure on a given day is the sum of new exposures over the diary events of that day. However, running exposures should not be summed across diary events; SHEDS outputs the time series and calculates a time averaged value from the time series on a daily basis and over the whole simulation. The average running exposure on a particular day is the duration-weighted average of the running exposure on the diary events of that day.

The reason for maintaining separate totals for the dermal-hands and dermal-body exposure surfaces is due to the possibilities of hand washing and hand-to-mouth transfer. Both reduce hand loading without altering the body loading. Once the loading on the hands reaches zero, no more can be transferred to the mouth, though body loading might not be zero.

The three types of contact media (air, matter, and surface) require three different forms for the exposure equations, in part because the concentrations are in different units in each case: surfaces use micrograms per square centimeter ($\mu g/cm^2$), air uses micrograms per cubic meter ($\mu g/m^3$), and matter (dust or soil) uses micrograms per gram ($\mu g/g$). Only inhalation is relevant for air. For surfaces and matter the chemical may either be dermally contacted and later ingested, or it can be ingested directly. The user selects which option is operative (see SHEDS-Multimedia version 3 User Guide). SHEDS version 3 does not include exposures to applicators themselves that may be incurred during the application process; all exposures are considered to be "post-application."

2.6.2 New Exposure

2.6.2.1 Inhalation exposure

Equation 2-3

$$\mathbf{E_{i,e} = C_{a,e} * BVR_e * METS_e * T_e}$$

where

$E_{i,e}$ = inhalation exposure for a given diary event *e* [µg]
$C_{a,e}$ = airborne concentration of chemical for the diary location [µg / m^3 air]
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]

The METS values are sampled from activity-specific distributions, with a new sample for every diary event. The distributions themselves are listed on the input file "diarymets," which 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. (2007). 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.

SHEDS assumes no carryover of inhalation exposure from one event to another. This seems reasonable because the processes of absorption into the blood or removal from the lungs are fast compared with diary event durations. The amount absorbed during each event is a fraction of $E_{i,e}$, and it

US EPA ARCHIVE DOCUMENT

is assumed that the remainder is exhaled. Thus, new exposure and running exposure are equivalent for the lungs.

2.6.2.2 Dermal exposure to surface residues (transfer efficiency method)

The SHEDS user has two options for calculating dermal residue exposure: the transfer efficiency method and the transfer coefficient method. A dermal transfer coefficient is a lumped parameter that aggregates mass transfer associated with a series of dermal contacts with a contaminated medium (units of area per time); a dermal transfer efficiency is a unitless fraction of chemical transferred from an object or hand to the mouth (see Cohen Hubal et al., 2000). Dermal transfer coefficients 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). They can also be calculated by multiplying a dermal transfer efficiency (unitless) by skin surface area contacted per time (cm^2/hr) (Zartarian et al., 2005). Dermal transfer coefficients and transfer efficiencies are both multiplied by a surface residue loading (mass per area) to model dermal exposure. Different surface residue collection devices are available (e.g., aluminum plates, rollers, sleds, hand presses, hand wipes, surface wipes); some collect dislodgeable residues and some collect total residues. These methods, as well as current methods for obtaining dermal transfer efficiencies and dermal transfer coefficients, have inherent uncertainties. Thus, it is important to consider matching the correct type of surface loading with the corresponding transfer factor (i.e., transfer efficiency or transfer coefficient) when developing inputs for modeling dermal exposure.

A given model run must utilize either the transfer coefficient or the transfer efficiency approach, as decided by the user. The same approach must be used for all surfaces. The current version of the SHEDS code requires that all exposure pathways use the same approach, either transfer coefficients or transfer efficiencies. Users of transfer coefficients might not have a readily available distribution applicable to the mouthing of objects. In that case, one can estimate an appropriate mouthing transfer coefficient by folding together distributions for object-mouth contact frequency, the object-mouth contact area, and the object-mouth transfer efficiency. The product of these three variables is effectively the mouthing transfer coefficient.

Equation 2-4

$$\mathbf{E_{r,j,e} = SR_e * A_{j,e} * CR_{j,e} * TE_e * T_e}$$

where

$E_{r,j,e}$ = new dermal residue exposure for a given diary event *e* for body part j [µg]
SR_e = available loading of chemical on the surface contacted [µg / cm^2]
$A_{j,e}$ = surface area of body part j (uncovered) [cm^2]
$CR_{j,e}$ = surface contact rate for body part j [1/hr]
TE_e = transfer efficiency [-]
T_e = duration of diary event [hr]

Here CR 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. See the equation below for object-to-mouth transfer.

2.6.2.3 Exposure to surface residues (transfer coefficient method)

Equation 2-5

$$\mathbf{E_{r,j,e} = SR_e * TC_{j,e} * T_e * Ad_{j,e}}$$

where

$E_{r,j,e}$ = new dermal residue exposure for a given diary event *e* for body part j [µg / event]
SR_e = available loading of chemical on the surface contacted [µg / cm^2]
$TC_{j,e}$ = dermal transfer coefficient for body part j [cm^2 / hr]
T_e = duration of diary event [hr / event]
$Ad_{j,e}$ = adjustment factor for body part j: for 'hands' there is no adjustment; for 'body' this is f_uncloth; the fraction of the body that is unclothed; and for 'mouth' it is bioavR, the bioavailability of residues.

The body part 'j' may be hands, body, or mouth. 'Hands' refers to the sum for both hands, 'body' refers to all dermal exposure except the hands, and 'mouth' refers to the direct mouthing of contaminated surfaces such as toys.

2.6.2.4 Dermal exposure to dust/soil

Equation 2-6

$$\mathbf{E_{m,j,e} = C_{s,e} * A_{j,e} * CR_{j,e} * Adh * T_e}$$

where

$E_{m,j,e}$ = new dermal dust/soil exposure for a given diary event *e* [µg]
$C_{s,e}$ = concentration of chemical in dust/soil [µg/g]
$A_{j,e}$ = surface area of body part j [cm^2]
$CR_{j,e}$ = surface contact rate for body part j [1/hr]
Adh = soil-skin adherence factor [mg/cm^2 x 1g/1000 mg]
T_e = event duration [hr].

Dust and soil concentrations are measured in units of micrograms of the target chemical per gram of matter. The dermal adherence is input to the model in units of milligrams per square centimeter of skin, so for units consistency the adherence is divided by 1000 in the code (to produce g/cm^2). The same adherence factor applies to both hands and body. Note that the exposure measures the amount of the chemical agent in contact with the skin, not the total amount of dust or soil. The same equations are used for both dust and soil contact.

2.6.2.5 Direct soil/dust ingestion

Two methods are available in SHEDS for determining soil and dust ingestion. In SHEDS, the terms “soil” and “dust” refer to the location of the material rather than to any physical or chemical properties. There are two distributions for direct ingestion, one applies to dust (for indoor events) and the other to soil (for outdoor events). Both distributions have units of (mg/hour) and are each sampled once per person. The soil ingestion rate is applied to all outdoor events, but there will be no new exposure unless the person happens to be in contact with an affected contact medium (the only one currently in

SHEDS is Ysoil). Indoors, the dust ingestion rate applies to all events except sleeping. Again, no new exposure will result unless the person contacts an affected medium (Tdust or Udust).

SHEDS does not actually track total soil and/or dust ingestion; it only tracks the amount of chemical agent ingested via this pathway. When not in contact with an affected medium, the amount ingested is recorded as zero. This is not meant to imply that the person ingested no dust/soil. The new exposure resulting from soil/dust ingestion is

Equation 2-7

$$\mathbf{E_{m,\,g,\,e} = C_{s,\,e} * Ing_{e} * BioavM * T_{e}}$$

where

$E_{m,\,g,\,e}$ = new dust/soil exposure for a given diary event *e* for GI tract [µg]
$C_{s,\,e}$ = concentration of chemical in dust/soil [µg/g]
Ing_{e} = dust/soil ingestion rate [g/hr]
BioavM = bioavailability for chemical in dust/soil [-]
T_{e} = event duration [hr].

The factor "BioavM" is the fraction of the ingested chemical mass that is in a form capable of being extracted and absorbed through the GI tract lining. In SHEDS, GI tract absorption is not instantaneous, but is modeled like dermal absorption. A fraction (proportional to event duration) of the GI tract loading is absorbed on each diary event, until GI tract voiding occurs. Whether this fraction should be applied to the exposure or to the dose calculation depends on the definition of the agent of the GI tract exposure. If the agent is defined narrowly (that is, as the amount present in a form that can be absorbed) but the total chemical mass is being tracked when it is outside the body, then the bioavailability fraction should be set to a value below 1. Otherwise, the bioavailability should be set to 1. The user should ensure that the GI tract absorption rates are consistent with the selected interpretation of the bioavailability term.

2.6.2.6 Indirect soil/dust ingestion

This is the alternate approach for determining the ingestion of chemical carried by dust or soil. It is based on the dermal hand loading of dust/soil, and uses the same hand-to-mouth transfer logic as is used for the transfer of residues. This equation is always used for residue transfer, but is only used for soil/dust transfer if the indirect method was selected. Some users prefer to use the direct soil/dust ingestion method because they have better knowledge of the input variable distributions involved. In that case, the hand-to-mouth transfer for soil/dust is disabled, to prevent double counting of the soil/dust ingestion.

2.6.2.7 Hand-to-mouth exposure transfer (residues, and optionally for dust/soil)

Equation 2-8

$$\mathbf{E_{hm,\,e} = (DE_{h,\,e} / 2) * HF_{e} * [1 - (1\text{-}MRE)\text{^}(FQH_{e} * T_{e})] * bioav}$$

where

$E_{hm,\,e}$ = new hand-to-mouth ingestion exposure for event *e* [µg]
$DE_{h,\,e}$ = running dermal exposure (hand loading) for diary event *e* [µg]

HF_e = fraction of one hand that enters the mouth [-]
MRE = mouthing removal efficiency (fraction transferred to mouth) [-]
FQH_e = frequency of hand-mouth activity [mouthing events/hr]
T_e = duration of diary event [hr]
bioav = bioavailability for residues or dust/soil

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. 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 initial chemical mass on the part of the hand entering the mouth is simply the loading on one hand ($DE_{h,e}$ /2), times the fraction of the hand which enters the mouth (HF_e). The second part of the transfer equation (the term [1 - (1-MRE)^(FQH_e* T_e)]) represents the fraction of this mass that is transferred. The fraction transferred is calculated as follows. The variable FQH gives the frequency of hand mouthing events per hour. For a diary event of duration M minutes, the expected number of hand mouthings during this event is (FQH_e x M / 60), or (FQH_e x T_e), where T is the event duration in hours. As an example, suppose each hand-mouthing event results in removal of MRE = 20% of the available loading (that which enters the mouth). If one starts with a loading of 1000 (units are irrelevant for this example), then the first mouthing removes 200, leaving 800 behind. The second mouthing now involves a loading of 800, of which 20% (or 160) is removed, leaving 640 behind. The third mouthing results in 640x20% = 128 being removed, leaving 512 behind. It should be clear that every mouthing means that the amount left behind is multiplied by another factor of 0.80, or (1-MRE). After N mouthings, the amount left behind would be 1000 x 0.80^N. Using variables instead of constants, the fraction removed on N mouthings is (1-MRE)^N, and here N=(FQH_e x T_e). The fraction transferred to the mouth is one minus the fraction left behind, so the transfer fraction is [1 - (1-MRE)^(FQH_e *T_e)].

The scenario presented here for hand-mouthing would imply that the power (FQH_e *T_e) should logically be an integer. In practice, diary events have variable duration, and (FQH_e *T_e) will often have a non-integer value. 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 HF_e) and one partial mouthing (an area somewhat less than HF_e).

Note that at most, only an amount $DE_{h,e}$ /2 * HF_e can be removed by hand-to-mouth transfer during one diary event. For example, if HF_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.

'$Bioav_c$' is the bioavailability term. The bioavailability refers to the fraction of the ingested chemical mass that is in a form that makes it available for absorption through the GI tract lining into the blood. Whether this fraction should be applied to the exposure or to the dose calculation depends on the definition of the agent of the GI tract exposure. If the agent is defined narrowly (that is, as the amount present in a form that can be absorbed) but the total chemical mass is being tracked when it is outside the body, then the bioavailability fraction should be set to a value below 1. Otherwise, the bioavailability should be set to 1. The user should ensure that the GI tract absorption rates are consistent with the selected interpretation of the bioavailability term.

US EPA ARCHIVE DOCUMENT

The hand-to-mouth transfer equation is always used for surface residues, but is one of two options for calculating soil/dust ingestion. This is the indirect method, as opposed to direct soil ingestion. The indirect method still uses the above equation as a hand removal rate, but does not add this to the GI tract loading, as that would result in a "double counting" of soil/dust ingestion.

2.6.2.8 Object-to-mouth exposure transfer (transfer efficiency method)

Equation 2-9

$$\mathbf{E_{om,\,e} = C_{surf,e} * object_ratio * om_area * [1 - (1- om_transfer)^{(om_freq_e * T_e)}] * bioavr}$$

where

$E_{om,\,e}$ = new object-to-mouth ingestion exposure for event *e* for residues [µg]
$C_{surf,\,e}$ = chemical residue concentration on contacted surface for event *e* [µg/cm^2],
object_ratio = concentration ratio for object to nearby surface [-]
om_area = area of the part of the object that enters the mouth [cm^2]
om_transfer = fraction transferred to mouth from object surface per mouthing event [-]
om_freq_e = frequency of hand-mouth activity for event *e* [mouthing events/hr]
T_e = duration of diary event [hr]
bioavr = bioavailability in the GI tract for residues [-]

The object-to-mouth transfer uses exactly the same logic as the hand-to-mouth transfer presented above. The only differences are in the names of the variables. 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 "object_ratio", which gives the typical concentration ratio for toy surfaces to room surfaces. The variable $C_{surf,\,e}$ will be event-dependent, so for example if a child is contacting the 'Thard' surfaces during some diary event, then the toys the child plays with at that time are with are assumed to have concentrations proportional to the Thard concentration. When the child plays in the untreated area, the toys are likely to be cleaner. This effectively implies that the toys are not moved from place to place with great frequency.

If the dermal transfer coefficient approach is used instead of the dermal transfer efficiency approach, then object-to-mouth transfer is calculated using the same equation as for hand and body dermal contacts, but with a different transfer coefficient.

2.6.3 SHEDS Exposure Adjustments

In SHEDS the running exposure is the current mass of the agent chemical being carried on the exposure surface. For example, the running dermal exposure is the mass of chemical (µg) on the skin, summed over both hands. Mass units are used in preference to others such as mass per unit area or mass per unit body weight, to facilitate conservation of mass during the acquisition and removal processes.

The exposure undergoes a series of adjustments on each diary event. It starts with the carryover exposure from the prior diary event. New exposure is then added. The totals are then checked against the maximum dermal loading limits. Next SHEDS implements absorption or other removal processes, which move some of the chemical from the given body part into the blood. Then hand-to-mouth transfer is calculated. Finally, removal processes such as washing and bathing are simulated.

The adjustments for each diary event are estimated and are made "instantaneously." For example, all new exposure occurs before any absorption, and all absorption occurs before any hand-to-mouth transfer, and so on. While these processes ought to be ongoing and simultaneous, using discrete time steps in modeling is a well-established practice that should be reasonable unless the diary events become too lengthy.

Not all of the SHEDS adjustments to running exposure affect all seven pathways. For example, the carryover loading does not apply to the lungs. SHEDS assumes that any chemical not immediately absorbed into the blood from the lungs will be exhaled, so the lungs start every diary event with zero running exposure. Also, at 6 a.m. each day the GI tract is voided and the running exposures set to 0.

2.6.3.1 New exposure

As discussed above, the "new exposure" in SHEDS 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 nonzero concentration. Two methods are available in SHEDS for determining new exposure: either transfer coefficients or transfer efficiencies. The choice is made at the start of the model run and applies thereafter; the two approaches cannot be mixed. Transfer coefficients are expressed in units of (cm^2/hr). There are three transfer coefficients in SHEDS, one for the hands, one for the rest of the body, and one for the mouthing of objects (other than the hands). The transfer coefficients are multiplied by the event duration (in hours) and the surface residue concentrations (in $\mu g/cm^2$) to give the new exposures in (μg).

The transfer efficiency approach calculates new exposure as the product of a fractional dermal contact rate (1/hr), the event duration (hr), the available (unclothed) area of the body part (cm^2), the surface concentration ($\mu g/cm^2$), and a surface-skin transfer efficiency (no units). For the mouthing of objects, an equation analogous to the hand-mouthing equation is used. The inhalation exposure is the same regardless of whether transfer coefficients or transfer efficiencies are used; it is given by the product of the METS level (no units), the basal ventilation rate (m^3/hr), the event duration (hr), and the air concentration ($\mu g/m^3$). The METS levels are part of the CHAD system and they represent the ratio of activity-specific metabolic rates to basal metabolic rates. These ratios are more stable across the population than the activity-specific metabolic rates are alone.

2.6.3.2 Maximum dermal loading

This check ensures that the dermal loading does not exceed some specified upper limit. For this purpose, the combined loading for both residues and matter (dust/soil) is examined. As a model input, the bound is given as maximum loading per unit skin area (micrograms of chemical per square centimeter of skin). Separate distributions may be entered for the hands and for the rest of the body, although generally one might expect them to be the same. For checking the dermal loading, the maximum dermal loading is multiplied by the hand skin area to obtain a maximum permitted mass on the hands; and similarly by the unclothed body skin area to obtain a maximum mass on the body. If the sum of the dermal loadings exceeds the maximum dermal loading, both components (residue and matter) are reduced proportionally until the limit is reached. If the actual loading is below the limit, no adjustment is made.

2.6.3.3 Absorption

On each diary event, the amounts absorbed into the blood are determined and removed from the exposures (preventing absorbing them again later). For the lungs, the absorption is a simple fraction of the amount inhaled; anything not absorbed is considered to be exhaled and is not available as carryover exposure to later events. Note that the time scale for pulmonary absorption is relatively fast compared to the diary event duration. For the other types of exposure, the absorption rates are specified on the input files as fractions per day. These are multiplied by the diary event duration (converted to days) to determine the fraction absorbed. The model assumes that the time scale for systemic absorption into the blood from the skin is slow compared to diary events, so this fraction is never close to one for a single diary event. Due to the smallness of this fraction per event, the amount absorbed is considered to be directly proportional to the diary event duration, without any corrections for non-linear effects.

US EPA ARCHIVE DOCUMENT

2.6.3.4 Hand-to-mouth transfer

Hand-to-mouth transfer of loading is both a source of new exposure for the GI tract and a removal process for the hands. In SHEDS it is determined after the absorption but before other removal processes. If the hand-to-mouth transfer calculations were positioned with other forms of new exposure, then the potential would exist for supra-maximal loadings to occur and be transferred to the mouth before the maximum dermal loading check could be made. This would be incorrect, as the maximum dermal loading should apply instantaneously. That is, if the new exposure results in excessive loading, then it essentially does not stick to skin at all but just falls off immediately, before there is time to transfer it to the mouth.

Given that the hand-to-mouth transfer is a dermal removal process, the question arises whether it should be considered before, during, or after the other removal processes. The most relevant competing removal process is hand washing. When children play and get their hands dirty, it is likely that they have enough time to mouth their fingers before their hands can be washed. Hence, the hand-to-mouth transfer should be evaluated before the hand loading is reduced by washing.

The equation for hand-to-mouth transfer is discussed in detail in the section on new exposures above; it is presented there because it amounts to a form of new exposure for the GI tract, and it also serves as a template for discussing the object-to-mouth exposure equation.

2.6.3.5 Other removal processes

There are three other removal processes for dermal loadings allowed in SHEDS that have not yet been discussed. Bathing (including showers and swimming) removes a fraction of the dermal loading from both the hands and the body. Hand washing removes a fraction from the hands only and does not affect the rest of the body. The third removal process is called “dry removal” or “brush-off,” which removes a fraction of the loading on all diary events, not just those involving water.

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. The consequence is that SHEDS diaries may sometimes indicate baths, but other diaries may have long stretches without any explicit mention of them. Each individual 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 ‘remv_bath,’ applied to both the body and the hands.

Hand washing is never explicitly indicated in CHAD and must be added by the model. Each person is assigned a value for the ‘washprob’ variable drawn from a distribution, which indicates the mean number of hand washings per day for that individual. It is assumed that the hourly probability for hand washing is 1/16 of this daily value, while awake. Thus, if a person had a daily mean of 4.0 hand washes, then the likelihood of hand washing during any waking hour is taken to be 25%. On diary events shorter than one hour, the probability is reduced proportionally to the event duration. In this example, a 12 minute event would have a 5% hand washing probability, since it lasts 1/5 of an hour. The efficacy of removal for hand washing is given by the ‘remv_wash’ variable.

This methodology for hand washing results in a variable number of washes from day to day. By the same token, the hand washes occur at different times each day, even when the same CHAD diary is used for the activities. Since the hourly probabilities are never integers, there is no need or advantage in requiring the daily mean to be an integer either. A person could average 1.6 hand washings per day, giving a 10% hourly probability. Therefore, the ‘washprob’ variable can be sampled from a continuous distribution, without needing to be restricted to integers.

The dry removal process can occur during all events other than bathing, washing, or sleeping. Its rate is set by the user-specified variable 'remv_dry,' which gives the removal fraction per hour. It represents the "rubbing off" of dermal loading that occurs during normal waking activity. If the user does not wish to utilize this process, the rate constant should be set to zero.

Multiple removal processes do not occur on the same diary event. Bathing has the highest priority. Hand washing can occur only if bathing does not. Dry removal only occurs if neither bathing nor hand washing are indicated.

2.7 Calculation of Dose Time Series for the Individual

Absorbed dose is defined in SHEDS-Multimedia as the amount of chemical that enters the human target after crossing the exposure surfaces (skin of hands, skin of rest of body, GI tract, lung). Blood dose is the mass of chemical or metabolite in the blood. SHEDS is primarily an exposure model, but it can also estimate dose using a simple built-in pharmacokinetic (PK) model. The SHEDS exposure profiles also can be exported for use in more advanced external PK or PBPK (physiologically-based pharmacokinetic) models to estimate dose. Once the dose profiles are obtained, they can be summed across routes and pathways to yield an individual's aggregate dose profile for the chemical of interest.

As discussed above, each exposure surface has exposure masses that are adjusted on each diary event. One of the adjustments is a subtraction for the amount absorbed. These absorptions lead to increases in the blood dose. SHEDS can report the absorbed dose by body part and the total absorbed dose for each diary event. Note that this absorption may occur several events after the chemical was originally contacted. Therefore, one may find dose increases even during events that take place away from new sources.

The amount of the absorption for each pathway is determined by the exposures and the relevant absorption variable, but the absorption does not directly depend on the details of the built-in PK dose model. The PK model in SHEDS is applied only after the chemical is absorbed into the blood. It is assumed that all of the absorbed chemical enters the blood and immediately is metabolized into another main product (one metabolite only) plus some side products that are disregarded. The mass ratio of the metabolite to the parent chemical, converted on a molar basis, is provided as a model input. If no metabolism occurs for the selected target chemical, the user should set this ratio to one. In that case, the metabolite should be interpreted as referring to the parent chemical. On each diary event, the absorbed metabolite masses are added to the blood dose, while a first-order elimination process is utilized for elimination of the metabolite. This elimination represents kidney filtration and possibly liver metabolism, rather than urinary excretion directly. Its rate is given by the input variable 'elimr_blood,' measured as a fraction per day.

The absorbed dose for each pathway, multiplied by the metabolite mass ratio, is added to the blood dose on each diary event. The seven pathway-specific dose variables are: 'dosHR' for hand residues, 'dosHM' for hand soil/dust, 'dosBR' for body residues, 'dosBM' for body soil/dust, 'dosGR' for GI tract residues, 'dosGM' for GI tract soil/dust, and 'dosLA' for lung inhalation.

On each diary event, the amount in the blood is reduced by kidney filtration. The rate variable is called 'elimr_blood' and is measured in units of [1/hr]. This variable is multiplied by the event duration in hours to determine the fraction eliminated. In the absence of new exposure, any amount in the blood will decrease approximately as a first-order exponential decay. Both the blood dose and the amount eliminated are expressed by pathway of entry, on each diary event.

The blood dose is essentially a running total or balance between two competing processes: the additions of chemical from absorption and subtractions of chemical by elimination. The dose will generally be nonzero at all times after the initial exposure occurs. It is not correct to sum the doses over

diary events. Instead, the proper measure of daily average blood dose is found by taking a duration-weighted average of the event-specific doses. This quantity is strongly impacted by the elimination rate. For some purposes, researchers are more interested in the total daily new dose (the sum of the new doses incurred during each diary event over the day; 'abs' variables in the SHEDS code). The 'dos' variables in the SHEDS code measure the running dose, which is the amount of chemical in the blood at the beginning of each diary event.

2.8 Extract Daily Statistics from Exposure or Dose Time Series

The preceding sections detail how SHEDS calculates time series for a set of exposure and dose variables. These represent a large amount of data, typically 15,000 to 20,000 events in a year-long simulation, for seven pathways, with five quantities ('exp' for new exposure, 'ldg' for running exposure, 'abs' for new dose, 'dos' for running dose, and 'elm' for elimination) reported for each pathway. In addition, date, time, and other variables must be retained as well. There are 10 megabytes or more of output data generated for each individual. Since a model run may generate thousands of individuals, this would amount to many gigabytes of output.

The amount of output data is reduced by retaining only certain summary data, while overwriting the most detailed data sets with the ones generated for the next individual. Even so, SHEDS retains about ½ megabyte per person for a one-year simulation period, if the option for retaining all variables is selected (this option is necessary if sensitivity analysis is desired).

The primary aggregation level produces daily summary statistics. The exp, abs, and elm variables represent changes per diary event, and are summed to obtain daily totals. The ldg and dos variables represent accumulated masses, for which daily values are obtained by taking duration-weighted averages of the event-specific values.

Two daily output summary files are constructed, differing only in their units. The file "allpersondays_ug" gives all the daily totals for exposure and dose in terms of micrograms (for ldg and dos variables), or micrograms per day (for exp, abs, and elm variables), for each person. The filename suffix "_ug" indicates micrograms, whereas the filename suffix "_mgkg" indicates units of (mg/kg/day). Hence, "allpersondays_mgkg" contains the same list of variables as allpersondays_ug, only using (mg/kg/day) for the exp, abs, and elm variables. Technically, the ldg and dos variables are in (mg/kg). To convert from (µg/day) to (mg/kg/day) the values are divided by 1000 to convert micrograms to milligrams, and divided by the body mass of the individual in kilograms.

Even the "allpersondays" files may become quite large. A standard variability run of 3000 persons for one year will have over one million records with 73 variables on each record. This is much too large to import into Excel, for example. Further aggregation is generally necessary before the data can readily be examined. The "dailymeans_ug" and "dailymeans_mgkg" files contain one record per person, with all the variables from the allpersondays files averaged across days for each person.

Finally, the "allvardaily_ug" and allvardaily_mgkg" files give the average values for each variable across all persons, for each uncertainty loop. In a variability run, the "allvardaily" files contain only one record each, giving the mean of each variable averaged across all persons and all days.

2.9 Generating Population Variability Estimates of Exposure or Dose

The steps described above are applied to each simulated individual. To obtain population estimates SHEDS-Multimedia 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. The chemical usage groups require the user to modify the application scenario probabilities appropriately, and (if necessary) the 'has_lawn', 'has_garden', or 'has_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-Multimedia 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 illustrative 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.

As with exposures, the built-in PK model in SHEDS produces time series of the dose for each route, 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. Then variability distributions (CDFs) of this metric across the selected population can be constructed. Or one could 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" on page 63.

3 MODEL INPUTS

SHEDS-Multimedia is a very flexible model which may be applied to a wide range of chemical exposure scenarios. To target 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

US EPA ARCHIVE DOCUMENT

characteristics, and various exposure and dose factors. Many of these inputs allow random sampling from probability distributions, although point values can also be used to fix certain inputs.

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

- Chemical Use-Related
  - Probability of using target chemical (on lawn, garden, pet, or indoors)
  - Pr (# applications per year) for each application type
  - Pr (month of application)
  - Pr (day of week application)
  - Pr (hr of application)
  - Probability of having a lawn, garden, pet
  - Pr (re-entry time indoors)
  - Pr (re-entry time outdoors)
- Activity-Related
  - Contact probability for various potential contact media
  - Probability of being in treated room while in home and awake
  - Max # days between baths
  - Soil/dust ingestion rate
  - Mean # hand washes per day per person
  - Fraction body unclothed
- Chemical Concentration/Residue-Related
  - Background and initial concentration/residues (Smooth and textured surfaces, dust, air, pet, lawn, garden, treated and untreated areas)
  - Decay rates (fraction lost per day)
  - Ratios for untreated to treated room concentrations
  - Object-surface concentration ratio
  - removal efficiency during bath/shower
  - removal efficiency during hand mouthing
  - removal efficiency during hand washing
  - residue-skin transfer efficiency
  - object-mouth transfer efficiency
- Non-Chemical Specific Exposure Factors
  - Body weight and surface area of hands and body
  - Dermal transfer coefficient for hands and body
  - Obj-mouth TE
  - soil-skin adherence factor
  - maximum dermal loading
  - Fraction of hand surface area that enters mouth
  - Obj-mouth contact SA
  - Body-surface and hand-surface fractional contact rate
  - object-mouth contact frequency
  - hand mouthing events per hour

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. Note that values in the demonstration file for SHEDS version 3 should NOT be assumed to be appropriate for actual modeling scenarios; they are for demonstration purposes only.

There are three types of stochastic sampling in SHEDS, two of which are under control of the user. For variables with a limited set of discrete output values, probability vectors are used. For variables that can take on any real-numbered value (or any value in a bounded range), continuous probability distributions are used. The third type is not under the control of the user; these involve hard-coded logic in the program code. An example of this last type is the selection of specific activity diaries from the diary pools.

3.1.1 Probability vectors

Certain input variables in SHEDS are restricted to discrete values or outcomes. To align the discussion here with the implementation in the code, a distinction is made between Bernoulli variables (with just two possible outcomes) and multinomial variables (with more than two possible outcomes). The latter are referred to here as "probability vectors."

For multinomial variables, the user supplies a list or vector of numbers that represent the probability for each outcome. For example, the vector (0.3, 0, 0.5, 0.2) would mean a 30% chance for the first outcome, no chance for the second, a 50% probability for the third, and a 20% chance for the fourth. The numbers in a probability vector must sum to one.

The interpretation of these outcomes depends on the variable in question. The probability vector as a type does not have a maximum length; any number of outcomes are permitted. However, for each input variable there are only a set number of outcomes that have meaningful interpretations, so there is a variable-dependent maximum length for the vector.

For Bernoulli variables (which may be interpreted as yes/no tests), the user simply furnishes a point value probability for one outcome, and the other outcome is automatically assigned if the test fails. For example, the variable "Do you have a pet?" has a point value for a "yes" outcome, rather than a probability vector.

In the SHEDS interface, the length of probability vectors is pre-set, and a check is made that the probabilities sum to one. If SHEDS is run in batch mode, the user must apply these checks. Many of the probability vectors in SHEDS refer to chemical usage, especially when the model is asked to generate the application dates. For each application type, probability vectors are assigned for "Day of Week," "Month of Year," and "Number of Applications per year." If the model is run with 4 application types, then 12 probability vectors are needed for these aspects alone. The five other probability vectors in standard use are listed below. In addition, the SHEDS model constructs probability vectors internally for the determination of contact with the various contaminated media. The user enters the individual probabilities that are used to construct these last vectors (not the vectors themselves), because the list of possible contact media is quite fluid in SHEDS and depends on other inputs or probability tests. For example, the pet is not a potential contact medium for persons without a pet, but it can be for other persons.

The full list of input variables for which the user must explicitly supply probability vectors is given in Table 3-2.

Table 3-2. List of user-supplied probability vectors.

Variable	Purpose
NumApps	Probability of exactly 1, 2, 3, … applications of same type at 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 application type used in the model run.
Month	Fraction of applications of this type occurring in each 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 type 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 application type used in the model run.
DayOfWeek	Fraction of applications of this type 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.
Reentry(outdoors)	Immediately after any outdoor application, people may be prohibited from entering the treated area. This is called the reentry prohibition. There are 9 categories: 0 hours (no prohibition), 1, 2, 3-4, 5-6, 7-12, 13-24, 25-48, and 49-96 hours. Each category has an overall probability for selection that the user may alter. Within each category, the possible outcomes are all equally likely and cannot be altered individually. For example, if the 5-6 category is selected, then the actual reentry time is equally likely to be 5 hours or 6 hours (50% chance for each). There is one vector for all outdoor application types, although it is sampled again for each application. The actual reentry time is an integer number of hours between 0 and 96 inclusive.
Reentry (indoors)	The same as above, but applied to all indoor applications.
TimeOfDay(outdoors)	The hour of the day when an application occurs. To keep the length of the vector and the randomly-generated application times reasonable, the hours are restricted to 7 a.m. to 6 p.m. The same vector is used for all outdoor application types.
TimeOfDay(indoors)	The same as above, but applied to all indoor applications.
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.

3.1.2 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 input 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
Point	v1	value	[-]	[-]	[-]
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 point and 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.

Not included in this table are bounds or truncation limits. These forms may be truncated at either or both ends. If a sample is generated outside the truncation limits, the value is shifted to the nearest limit. The third aspect relevant to most input variable distributions is their sampling frequency. Each input is sampled once for each simulated person, at least. But it is possible to request more frequent sampling. This is described in more detail below.

Note that statistical properties of distributions such as the mean and standard deviation always refer to the untruncated version. For example, if a normal distribution has part of its lower tail truncated, then the distribution of the values actually realized will have a mean exceeding that of the original normal, but have a standard deviation less than that of the original normal. For most forms, including the normal, the magnitudes of these changes are not simple functions of the truncation point. In practice, the methods of generating these forms rely on first sampling from the untruncated version, then adjusting the values of any outlying results.

Most of the SHEDS input variables are sampled from the above distributions. Usually these inputs are sampled independently, although it is possible to request correlated inputs, as discussed below. A summary list of the continuous input variables follows. For a full discussion of all inputs see Appendix F.

US EPA ARCHIVE DOCUMENT

Table 3-4. List of continuous variables in SHEDS.

Variable	Purpose
absf_lung	absorption fraction for lungs
absr_dm	dermal absorption rate per day for dust or soil
absr_gm	GI tract absorption rate per day for dust or soil
absr_gr	GI tract absorption rate per day for surface residues
adherence	soil-skin adherence factor
bioavm	bioavailability fraction for dust/soil
bioavr	bioavailability fraction for surface residues
contactb	body-surface fractional contact rate
contacth	hand-surface fractional contact rate
dermaxb	maximum dermal loading for body
dermaxh	maximum dermal loading for hands
elimr_blood	elimination rate from the blood
f_uncloth	fraction of body unclothed
has_garden_p	probability of having a vegetable garden
has_lawn_p	probability of having a lawn
has_pet_p	probability of having a dog or cat
hm_fraction	fraction of surface of one hand that enters mouth
hm_freq	frequency of hand mouthing events per hour
ingestion_indoor	dust ingestion rate (indoor, direct only)
ingestion_outdoor	soil ingestion rate (outdoor, direct only)
metab_ratio	mass ratio (metabolite/pollutant)
om_area	object-mouth contact area
object_ratio	object-surface concentration ratio
om_freq	object-mouth contact rate
om_transfer	object-mouth transfer efficiency
remv_bath	removal efficiency during bath/shower
remv_dry	removal efficiency during events without water
remv_mouth	removal efficiency during mouthing
remv_wash	removal efficiency during hand washing
transfer_dermal	residue-skin transfer efficiency
washprob	mean # hand washes/day per person
tc_b	surface-skin transfer coefficient for body (unclothed)
tc_h	surface-skin transfer coefficient for hand
tc_m	transfer coefficient for object mouthing

In addition to the above, there are 168 distributions assigned for activity-dependent ventilation rate distributions (METS distributions) on the file "diarymets;" these are not alterable through the interface. There are also 18 distributions for contact probabilities. Finally, there are up to 208 distributions relating to the distribution of concentrations at various time intervals after an application, for each application type. This is active only if the "postapp" method of determining concentration is used. If the "decay/dispersion" method is used, the maximum number of such distributions is only 92. Depending on

US EPA ARCHIVE DOCUMENT

the application types selected, the number of inputs could be significantly smaller. For example, if the only application type is PetSpot then just two inputs are needed for the decay/dispersion method, apart from background concentrations.

3.1.3 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, hour, or activity diary event.

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 input directory, but this is recommended only for experienced SAS users. See Description of Residential Input Files.

3.1.4 Correlation of Inputs

SHEDS Multimedia version 3 permits randomly sampled variables to be generated with correlation. For instance, one might randomly sample both the transfer coefficient for the hands and the transfer coefficient for the body, but one might desire to have some statistical relationship between them. An extreme solution to this and similar problems would be to reduce them to one random variable, and calculate one from the other deterministically. But SHEDS' ability to generate correlated random variables renders this unnecessary.

For any pair of input exposure factors, dose factors, or application-related inputs that are sampled from continuous distributions, the user may specify a desired level of Spearman (rank) correlation. (Pearson and Kendall correlations are not implemented in SHEDS for detailed computational and statistical reasons.) The method uses a modified NORTA approach (Smith and Glen, in preparation), but the details are too lengthy to be described here. 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 form and also satisfy the correlation targets.

One cannot request correlations for any variables that use point distributions or probability vectors. Only the continuous non-point forms (uniform, normal, lognormal, triangular, exponential, beta, gamma, and Weibull) can be used. There is no restriction on sampling frequency. It is possible to request correlations between two or more variables with different sampling frequencies.

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 for this that is automatically performed in SHEDS. There is also a slightly more stringent restriction on the correlation 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 definite, a message is printed to the log during the model run.

SHEDS Multimedia version 3 handles correlation between the values of two or more input variables simultaneously. A possible extension to be available in version 4 would allow correlation within a single input variable at different simulation times (that is, autocorrelation in sample values). Methods for inducing correlation or conditional probabilities among probability vectors are also being explored for inclusion in version 4.

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. Such quantities can be 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 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. A few input variables do not conform to the SHEDS "standard" that masses should be in micrograms. For example, the soil adherence variable has units of (mg/cm^2), which is traditional in the literature. When necessary, the model internally converts such units into its 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 familiar with SAS should have little problem. It is strongly recommended that others employ the SHEDS interface.

US EPA ARCHIVE DOCUMENT

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 Demo File and multiple distributions for one variable

The SHEDS installation package and the user interface come with a demo file containing input distributions that do not reflect real world data. One feature of SHEDS is that a single continuous input variable can be assigned multiple input distributions. For example, the distributions could be age or gender specific. The SHEDS interface only partially supports this feature. Where the demo input file (called "distributions") has multiple cases for a particular variable, then the interface will allow the user to modify the distribution for each case separately. However, the interface does have the capability to alter the rules that distinguish these cases.

If the user wishes to create new or different rules for selecting between multiple distributions, then the "distributions" file must be edited appropriately in SAS, but outside of the user interface. The rules are written as valid SAS expressions such as may be used in IF statements. For example, the distribution for males is indicated by the condition (gender="M"), while the one for females is indicated by (gender="F"). While age and gender are the two most usual variables used in conditions, many others are possible.

Each such case appears on a separate record on the "distributions" file, sharing the same value for the variable named "variable." The user should ensure that the logical conditions cover all possible cases, or unpredictable results may occur. The user should be an experienced SAS programmer, since these conditions effectively become new SAS code. Errors may occur, and the interpretation and resolution of such errors can be difficult. For this reason, the option of defining rules for multiple distributions was not included in the interface.

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 Monday in 2007, but will be a Tuesday in 2008, 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.

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.

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. Otherwise, there would be unnaturally low concentrations early in the simulation period, regardless of the start date. SHEDS therefore checks for chemical usage at least 30 days prior to "day 1" of the simulation. Therefore, the output may indicate that the house was sprayed on "day -22" or even on "day -185". Checks are made back to the last January 1 that occurred at least 30 days prior to the start of the simulation period. 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 Chemical Usage-Related Inputs

3.3.1 Application-specific inputs

Under the TimeSeries option there are no application-specific inputs. But when SHEDS is run using either the ModelDates or UserDates option, the model must set the dates, times, and amounts of chemical usage. These usages are called "applications," and directly affect the determination of the chemical concentration time series. SHEDS uses an absolute measure of time, not relative to the time of an application. While it is possible to have an application at the start of the simulation period, SHEDS permits delays in application, or multiple applications, or no application at all for some persons.

There are 10 application types currently defined in SHEDS (Table 2-3). The user selects which of these are to be used in a given model run, from any single one up to all ten. For each application selected, certain information must be provided. If the UserDates option is used, then the dates and times for each application type are needed. More inputs are needed for ModelDates, since the model must be provided with the information for assigning dates that are random yet conform to specified usage patterns across the population.

3.3.1.1 Inputs for the UserDates option

For each selected application type, the user must enter the list of application 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 an application occurs seven days before the start of the simulation, another application on day zero which is the day before the start of the simulation, a third application 14 days after the start of the simulation, and a final one on day 90. A time of day, from 7 a.m. to 6 p.m. inclusive, is also entered for each application type. If this were set to noon, then all applications of this type would happen at noon, for all persons. Other application types could be set to different hours.

It is common in UserDates runs to apply the chemical at the start of the simulation. This can be achieved by setting the application date list to (1), but note that the first day would then include some time before the application actually occurs (since it cannot occur before 7 a.m.). 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.

In the UserDates option, the variables Has_lawn, Has_garden, and Has_pet are not relevant. Every simulated person shares the same application dates. Therefore, if a lawn application is selected, all persons are assumed to have a lawn. Similar remarks apply to the garden and pet, as well.

3.3.1.2 Inputs for the ModelDates option

The ModelDates option considers the population variability in chemical usage. The

SHEDS user specifies the list of application types to be considered in a given SHEDS run. However, a given individual will probably not experience all of the selected application types, unless the target population is narrowly defined. The two lawn application types can only occur for those persons with Has_lawn=1. Similarly, the two pet application types can only occur for persons with Has_pet=1, and the vegetable garden application can only occur for persons with Has_garden=1.

For each application type the overall probability of use must be provided. This is only applied to those who meet the conditions mentioned above. For example, suppose the probability of having a lawn (the variable "Has_lawn_p") is 80%, and the probability of treating the lawn 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).

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

1. Pr (# applications per year)
2. Pr (month of application)
3. Pr (day of week application)

The number of applications per year has room for 12 probabilities in the interface; a longer list can be used in the SHEDS SAS code. The first or leftmost probability is the likelihood of having exactly one application of the given type in one year, given that at least one occurs. The second probability represents exactly two applications, given that at least one occurs, and so on up to 12 applications. The sum of the probabilities must equal one, since one of these outcomes must be chosen. For a simulation period less than a year, this probability vector still represents a full year. The model may select application dates before, during, of after the simulation period.

The probability vector for the month of application 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. 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.

Two other probability vectors are needed for the ModelDates option. These give the hour of the day when the application occurs; one vector for all indoor application types and the other for all outdoor application types. These vectors have 12 slots, with the first representing 7 a.m., the second 8 a.m., up to the twelfth slot representing 6 p.m.

3.3.1.3 Re-entry prohibition times

The re-entry prohibition period is also discussed in the Methods section. It represents a short period immediately after an application when people are not permitted to enter a treated area. There are two re-entry time probability vectors, one for indoor applications and the other for outdoor applications. Each vector has nine possible outcomes. These correspond to 0, 1, 2, 3-4, 5-6, 7-12, 13-24, 25-48, and 49-96 hours. In SHEDS, the re-entry time is selected separately for each new application. If the

outcome is a single number, then the prohibition lasts that many hours. If the outcome is a range, then the prohibition is picked uniformly from that range.

The actual time until first contact may be longer than the prohibition period. The activity diary must indicate the correct sub-micro (see glossary), and a contact probability check must be passed. If the application is outdoors and the person happens to have no outdoor time on his or her diary, then no contact or new exposure can occur until a day when a different diary is used.

3.3.2 Co-Occurrence Factors

When the ModelDates option is used, the default assumption is that the actual dates of chemical usage (applications) are selected independently. To change this assumption, the user should specify co-occurrence factors.

For a single application type, only a blackout period can be imposed. This prevents two or more applications of that type 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 no other application of the same type can occur between days 93 and 107 (exclusive).

When two or more application types are used in the same model run, with the ModelDates option, then two co-occurrence parameters may be specified. One is the multiplicative factor for adjusting probabilities on each nearby day, and the other is the number of days on each side to be adjusted. Figure 2-2 and Figure 2-3 illustrate these adjustments.

3.4 Activity Diary-Related Inputs

3.4.1 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 contains over 22,000 one-day (24 hour) activity diaries, consolidated from more than a dozen original studies. Other sources of activity diaries could potentially be used in SHEDS; 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. SHEDS permits the user to reduce this maximum, so that longer diary events are broken into shorter pieces. 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 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 by a SAS programmer, outside of the user interface.

3.4.2 Re-extracting data from CHAD for use in SHEDS

The SHEDS installation package already contains the CHAD data needed for SHEDS. If these data are not available or if CHAD is updated, the user may wish to re-extract the data directly from the CHAD database. To do this, first obtain the personal computer version of CHAD. The current version is "chad2000_v1047.mdb." Run this file and click on the large central button labeled "CHAD." On the next screen click on the "Select Filter" button in the upper right corner. Then without creating a filter (so that all records are retained), click on the "Send it to a file" button. Two screens will pop up, one for questionnaire data and the other for diary data. For the questionnaire data, leave all the settings at their defaults, except perhaps for the name of the file to be saved. The path should be changed to the \default directory for SHEDS, but the filename should otherwise not be changed.

For the diary data, only 5 variables should be kept. First click on the check box labeled "Select all variables" so that it is unchecked. Then check the boxes for Start Time, Duration, Activity, and Location. The CHADID box always remains checked. When only these five variables are checked, click on the Exit Door button. Shortly, a popup box will prompt the user for the destination filename. This should also be redirected to the SHEDS \default directory.

3.4.3 Pre-processing the CHAD files

Once the files "\default\questionaire_data.txt" and "\default\diary_data.txt" exist, the "%diarypreprocessing" macro in SHEDS needs to be run. This macro must also be run if the user wants to alter the rules for acceptance/rejection of the CHAD diaries. 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 two ".txt" files exported from CHAD, 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. 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.
sleep>0	There must be some sleep time.
LocUX<60	Less than 60 minutes in unknown locations.
ActUX<120	Less than 120 minutes in unknown activities.

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. Assume that the user has specified five micros: in_home, in_veh, in_oth, out_home, and out_oth (see Table 2-2). 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

US EPA ARCHIVE DOCUMENT

Variable	Meaning
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
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

US EPA ARCHIVE DOCUMENT

Variable	Meaning
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
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 can be found in the CHAD documentation.

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.

US EPA ARCHIVE DOCUMENT

3.5 Contact Probability-related Inputs

The contact probability-related inputs control the likelihood of the simulated person coming into contact with particular media. For CHAD diary events that take place inside the residence, three input variables are needed to determine the "sub-micro" or room type (treated, untreated, or neither). These are:

- Probability of being in treated room while in home and awake
- Probability of being in untreated room while in home and awake
- Probability of being far from a source while in home and awake

As discussed in the Methods section, the "treated room" is any room where the chemical has been applied (or will be applied). The "untreated room" is any room into which the chemical may disperse. The third category refers to locations with no possibility of new exposure.

While individuals are sleeping, SHEDS assigns them to an untreated room, and it is assumed that no dermal contact or soil/dust ingestion occurs while sleeping.

The above three variables may be assigned distributions. These may be point values, but this is not necessary. In a general population, some individuals will probably treat a larger fraction of their house than others. Correspondingly, the probability of being in a treated area should be larger for such persons. For each person, a sample is drawn from each distribution. The sum of the three values is determined, and each value is divided by the sum. The three adjusted values then sum to unity.

In addition, the following probabilities are needed to determine the specific media being contacted. In Table 3-7 these are grouped according to the probability vectors formed from them.

Table 3-7. Contact probability variables.

Group	Variable
1	Contact probability for smooth surface in treated room
	Contact probability for textured surface in treated room
	Contact probability for pet while in treated room
	Contact probability for no surface while in treated room
2	Contact probability for dust while in treated room
3	Contact probability for smooth surface in untreated room
	Contact probability for textured surface in untreated room
	Contact probability for pet while in untreated room
	Contact probability for no surface while in untreated room
4	Contact probability for dust while in untreated room
5	Contact probability for lawn while outdoors at home
	Contact probability for garden while outdoors at home
	Contact probability for pet while outdoors at home
	Contact probability for no surface while outdoors at home
6	Contact probability for soil while outdoors at home

The odd-numbered groups contain four probabilities each. As for the sub-micro probabilities, one value is drawn from each distribution, per person. Each value is divided by the sum of the four, to create a probability vector whose probabilities sum to unity. For dust and soil contacts, only one probability is given in each location. This is an example of a yes/no decision (either contact occurs on this diary event, or it does not). The probability of a "no contact" outcome for dust or soil is just one minus the probability of a "yes" outcome, so it does not need to be specified separately.

3.6 Concentration-Related Inputs

As discussed above, there are three options for specifying concentrations used in SHEDS. The following sections describe the relevant inputs for these different methods.

3.6.1 Decay-dispersion method

3.6.1.1 Indoor application types

For each combination of selected application type and directly affected contact medium there are three inputs for the decay-dispersion method. The indoor application types (InCCaer, InCCliq, InCarRm, InFIK, and InFog – see Table 2-3 for a description of these) each affect four contact media (Thard, Tsoft, Tdust, Tair). Suppose the user selects the indoor crack-crevice aerosol (InCCaer) application type. The distribution for the initial concentration resulting from an InCCAer application is needed for each of the four contact media in the treated room. The concentration units are ($\mu g/cm^2$) for the surfaces Thard and Tsoft, the units are ($\mu g/g$) for Tdust, and are ($\mu g/m^3$) for Tair. The decay rate for the concentration is needed in each of the four media, expressed as the fraction lost per day. The third variable is used to determine the rate of dispersion from the treated to the untreated room. This variable requires some explanation.

Suppose both the treated and untreated rooms start at zero concentration and an application occurs in the treated room. 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. Once the application has occurred, 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. There is also a particular ratio of concentrations between the untreated and treated rooms at this time.

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 assumes that the treated and untreated rooms are essentially equivalent in all respects except for the chemical concentration. In this case, there is a clear relationship between the flow rate, the time until the untreated room reaches maximum concentration, and the concentration ratio at this time (when the untreated room reaches its maximum). Specifically, the ratio of the fraction in the treated room lost to dispersion to the fraction lost to decay equals the ratio of concentrations in the untreated room to the treated room at the time of the untreated room maximum (assuming both these ratios are much smaller than one). The latter ratio is easily measured in practical situations, whereas the fraction lost to dispersion would be difficult to measure directly. Therefore, SHEDS requests as input the variable "UTratio," which is the ratio of concentrations in the untreated room to the treated room at the time of the untreated room maximum. SHEDS uses four such ratios, one for each contact medium.

3.6.1.2 Outdoor and pet application types

The two lawn and the one vegetable garden application type each affect one surface medium (Ylawn or Yveg) and the soil (the contact medium Ysoil). For each selected application type two distributions are needed, each medium requiring one for initial concentration and one for decay rate. 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.

There are two pet application types, each of which affects only one contact medium. For purposes of contact, the pet is effectively another surface medium. Unlike the other surface 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 selected application type affecting pets, a distribution for initial concentration and another for the decay rate are required.

3.6.2 Post-application method

For each combination of selected application type and affected contact medium there are four inputs for the post-application 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.

The indoor applications type affect eight media each (Thard, Tsoft, Tdust, Tair, Uhard, Usoft, Udust, and Uair). Thus, 32 input distributions are required for each selected application type. The outdoor application types affect two media each (one surface and one soil), requiring 8 distributions to be specified. The pet application types affect only one contact medium, thus requiring just four distributions each.

3.6.3 Time Series method

If the TimeSeries option for concentrations is used, SHEDS bypasses the application logic and expects the user to supply ready-made concentration time series, in SAS format. A sample concentration time series data set called “timeseries” is included in the \default directory.

The SHEDS user may create his or her 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 on page 25 (

Table 2-4).

In the previously mentioned example, there are three houses with two contact media defined in each. The contact media not appearing on this file effectively have zero concentration. The omitted records for July 19 and 20 may optionally be included on the input file with all the concentrations set to missing. The time series method does not currently support variation in concentration within each day.

As discussed earlier (where Table 2-4 is presented), the input data file is subset so that only those records for days within the overall simulation period are retained. This may lead to a reduction in the number of houses, as houses without any valid days are excluded. Note that SHEDS assigns a specific house to each person and all persons are sampled independently, so it is not possible to have the 'same' person exposed to the chemicals in more than one house, even if the data do not overlap in time.

If an input file is supplied that contains missing values (time gaps) within a house, the model will run but error messages will be generated on the SAS log. Persons assigned to that house will have invalid exposure estimates.

3.7 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.

3.7.1 Physical exposure factors

The potential for dermal exposure depends on the surface area of the two dermal compartments in SHEDS, 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 (up to age 20) and one for adults. For children, height is independent of weight and is normally distributed, with mean and standard deviation taken from the "bwsa" file. 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. However, a sample 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. NHANES III (http://www.cdc.gov/nchs/about/major/nhanes/datalink.htm#NHANESIII) 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 inhalation rate is related to the basal metabolic rate by

Equation 3-1

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 and another for all older persons. The skin surface area (SA) is measured in square centimeters (cm^2). Below age 6 the formula is

Equation 3-2

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

For those ages 6 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 point probability that applies to the entire target population. Each individual is then tested against these probabilities, for each of the three variables. There is variation across the population in the outcomes, even though the inputs are set to point values.

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 3 is not configured for lifetime simulations, nor for multi-year simulations of children.

3.7.2 Activity-related exposure factors

The activity-related exposure factors are variables describing aspect of personal behavior that directly affect exposure, excluding the usage-related factors and the contact probabilities. These variables include:

- maximum number of days between baths
- mean number of hand washes/day per person
- hand-surface contact rate

- body-surface contact rate
- hand-to-mouth contact rate (events per hour)
- object-to-mouth contact rate (events per hour)
- soil ingestion rate (outdoors)
- dust ingestion rate (indoors)
- fraction of surface of one hand that enters mouth

The purpose of the variable for maximum number of days between baths is described under "Removal Process" in "SHEDS Exposure Adjustments" in the Methods section. This is a probability vector of length seven. The first element is the probability of having a bath 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.

The mean number of hand washes/day per person is also described under "Removal Process" in "SHEDS Exposure Adjustments" in the Methods section. This mean number is considered to be a personal behavioral characteristic, and is sampled once per person. The actual number of hand washes changes from day to day, but the average over the simulation period should be close to the designated mean for the individual.

The hand-surface contact rate can be confusing, since it represents a fraction yet in some cases may exceed unity. Since skin area is already calculated for each individual separately under the physical exposure factors, the skin-surface contact area is determined by multiplying the skin surface area by the fractional contact rate per unit time. Suppose this rate were 5% per minute. That would imply all of the skin would contact the surface in 20 minutes, or 300% would contact the surface in one hour. The interpretation of these quantities is that multiple contacts by the same portion of the actual skin surface may occur, and are counted multiple times. A 300% result implies that all of the available skin surface averages 3 contacts per hour with the relevant contact medium. The hand-surface contact rate is re-sampled every hour.

The body-surface contact rate is similar to the hand-surface contact rate, except that it represents the fraction of the available (unclothed) body surface area, excluding the hands. Here, too, the value of this variable may exceed unity. The body-surface contact rate is re-sampled every hour. The hand-to-mouth contact rate gives the number of hand (full or partial) insertions per hour. This variable is re-sampled every hour. Since it is used as a rate constant and will be multiplied by the event duration, there is no necessity to restrict it to integer values. The object-mouth contact rate uses similar logic to the hand-to-mouth contact rate, except that it refers to toys and other objects instead of the hands.

The soil ingestion rate is used only if the direct soil ingestion method is selected. (The indirect method estimates soil ingestion from the soil loading on the hands and the hand-mouth contact rate.) The soil ingestion rate is measured in milligrams per hour, and is applied to outdoor events only. It is sampled once per person. The dust ingestion rate is similar to the soil ingestion rate, but is applied indoors (but not while sleeping). It is also applied while in vehicles, and is sampled once per person. A typical CHAD diary for a child has about 1-3 hours per day outside, about 10-11 hours sleeping, and about 11-12 hours indoors while awake (including roughly 1 hour in vehicles).

US EPA ARCHIVE DOCUMENT

The fraction of surface of one hand that enters mouth represents 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. This variable is sampled once per person.

3.7.3 Transfer-related exposure factors

The transfer-related exposure factors are variables that describe the tendency for the chemical to move from one compartment to another (generally on or off the human body). These variables include:

- soil-skin adherence factor
- maximum dermal loading
- object-mouth contact area
- object-surface concentration ratio
- fraction of body unclothed
- removal efficiency during hand washing
- removal efficiency during bath/shower
- surface residue to skin transfer efficiency
- removal efficiency during hand mouthing (saliva transfer efficiency)
- object-mouth residue transfer efficiency
- brush-off removal efficiency
- surface residue-to-skin dermal transfer coefficient for hand
- surface residue-to-skin dermal transfer coefficient for body
- transfer coefficient for object mouthing

The soil-skin adherence factor measures the tendency for the skin to accumulate soil loading while outdoors. In SHEDS, this variable is not dependent on event duration, but it is applied only when contact occurs. Therefore, when used it is multiplied by the hand-surface or body-surface contact rate and by the event duration. The units should therefore be interpreted as milligrams of soil per square centimeter of skin per contact. Note that the literature usually reports soil loading effectively as ($mg/cm^2/day$), as the loadings are often measured after accumulating over several hours of outdoor activity. Soil-skin adherence is sampled once per person.

The maximum dermal loading allows separate distributions for the hand and body. It is measured in units of micrograms of target chemical per square centimeter of skin. It is applied to the combined loading of residue, dust, and soil. When new exposure pushes the dermal loading over this limit, the loadings are reduced proportionally until they no longer exceed the maximum. The maximum dermal loadings are sampled once per person.

The object-mouth contact area is only used when the transfer efficiency option is selected by the user (see User Guide). This represents the effective surface area of toys and other objects that is 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. This variable is sampled once per person.

The object-surface concentration ratio is used to calculate the chemical concentration on the objects being mouthed. This is expressed as a ratio to the concentration on floor surfaces in the same location. It is unitless and is sampled once per person.

The fraction of body unclothed indicates the amount of skin available for contact with dust/soil, excluding the hands. It is sampled once per day.

The removal efficiency during hand washing is used to reduce the chemical loading on the hands during hand-washing events. This is not dependent on event duration. It is a unitless fraction, and is sampled once per hour.

The removal efficiency during bath/shower is applied during bath/shower/swim events. It is applied to both the hand loading and the body loading, and applied equally to residue, dust, and soil loadings. It is a unitless fraction, and is sampled once per hour.

The surface residue to skin transfer efficiency is used in one of two options (the other uses transfer coefficients) for calculating dermal exposure. It is unitless and is sampled once per hour. It represents the fraction of the dislodgeable surface residue that is transferred from a contact medium to the skin upon each contact.

The removal efficiency during hand mouthing (saliva transfer efficiency) is applied during hand-to-mouth contact events, on a per-contact basis. As discussed in the Methods section, a non-linear transfer equation is used when multiple contacts occur. It is unitless and is sampled once per hour.

The object-mouth residue transfer efficiency is used during object-to-mouth contact events. It is similar to the hand-to-mouth transfer efficiency discussed immediately above. It is unitless and is sampled once per hour.

The brush-off removal efficiency is 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, set this variable to a point value of zero. This variable is sampled once per person.

The surface residue-to-skin dermal transfer coefficient for hand is one of the three input variables that are used only when the transfer coefficients method for new exposure is selected by the user (see User Guide). Transfer coefficients are expressed in units of (cm^2/hr), and are simply multiplied by the relevant surface concentration to obtain the new exposure. All transfer coefficients are sampled once per hour.

The surface residue-to-skin dermal transfer coefficient for body is similar to the prior variable, except applied to the body instead of the hands. In SHEDS it is expected that the transfer coefficient is appropriate for an unclothed body, and that clothing will reduce the transfer proportionally. Thus, if the body is 20% unclothed, then the actual transfer to the skin is 20% of the value indicated by the transfer coefficient. Transfer to clothing is assumed to be ineffective in terms of causing exposure.

The transfer coefficient for object mouthing is used for the object-to-mouth contact pathway when the transfer coefficients method is selected. It is applied to the surface concentration on toys and other objects. Like the other transfer coefficients, it has units of (cm^2/hr) and is sampled once per hour.

3.8 Dose Factors

If the user opts to use the built-in SHEDS pharmacokinetic model, input values for several dose factors are required. All of these variable values are chemical dependent, so the user must ensure that appropriate values are set for the selected target chemical. The dose factors are:

US EPA ARCHIVE DOCUMENT

- bioavailability fraction for surface residues
- bioavailability fraction for dust/soil
- dermal absorption rate (fraction per day) for surface residues
- dermal absorption rate (fraction per day) for dust or soil
- GI tract absorption rate (fraction per day) for surface residues
- GI tract absorption rate (fraction per day) for dust or soil
- absorption fraction for lungs
- molecular weight mass ratio of the metabolite to the pollutant
- elimination rate (fraction per day) from the blood

The bioavailability fraction is defined here as the fraction of the amount ingested that is in a form available for absorption into the blood from the GI tract. There is one bioavailability fraction for surface residues and another for soil/dust. These are unitless and are sampled once per person.

The dermal absorption rates indicate the fraction of the dermal loading that would be absorbed on a daily basis. SHEDS calculates absorption on an event basis. Therefore, the fraction of the dermal loading absorbed per diary event is given by the product of the daily rate and the event duration (expressed as a fraction of a day). There are separate dermal absorption rates for residues and for dust/soil. The units are (1/day) and these variables are sampled once per person.

There are two GI tract absorption rates, which operate similarly to the dermal absorption rates. Here, too, the base rate is given as a fraction per day, and an event-duration-weighted fraction is absorbed every event. It is applied to the current GI tract loading on each event. The units are (1/day) and these variables are sampled once per person.

The absorption fraction for lungs differs from all the other absorption rates, since it is not duration dependent. It is unitless and represents the fraction of the inhaled volume that enters the blood. Any chemical not absorbed is assumed to be exhaled, so there is no carryover to the next diary event for the lung exposure. The absorption fraction for lungs is sampled once per person.

The molecular weight mass ratio of the metabolite to the pollutant is used for chemicals that quickly metabolize after entering the blood, but thereafter are slow to be removed. Thus, the main consequence of exposure is accumulation of a metabolite in the blood. If the target chemical does not metabolize (or is slow to do so), then it would be appropriate to set this input to a point value of one. In any case, this variable should be set to a point value since the same metabolite is formed by each person.

The elimination rate (fraction per day) from the blood is used to reduce the blood loading (called "dose" in SHEDS) of the target chemical, or more correctly its metabolite. Like the absorption variables, the fraction eliminated is given by the product of the daily rate and the event duration (expressed as a fraction of a day). This variable has units of (1/day) and is sampled once per person.

US EPA ARCHIVE DOCUMENT

4 OUTPUTS

4.1 Individual Time Series

Exposure time profiles for individuals are the basis of the SHEDS-Multimedia 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-Multimedia approach of tracking the sequence 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 PK or PBPK models to be used.

The output data set "diary" in the SAS work directory contains all the time series for the latest individual simulated in the model. Due to the large size of this data set (typically 3-10 megabytes per person), it is continually recycled within SHEDS and is overwritten by the data for the next person to be simulated. The time series for the key exposure and dose variables are aggregated to the daily level and stored permanently on the "allpersondays" data sets. There are two of these, one with the suffix "_ug" for results in units of micrograms, and the other with the suffix "_mgkg" which has results in units of (mg/kg/day). The data set 'diarycounts' contains the number of valid activity diaries for each cohort and day type. Note that the user can designate the folder (directory) where the output data sets are written.

4.2 Population Variability Distributions

One of the main purposes of SHEDS is to calculate the distribution of exposure across the target population. The model is designed so that each individual is randomly selected, but representative of the target 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 user must select some exposure or dose statistic of interest. Often, this is some time-averaged variable derived from the individual time series, such as the ADD (average daily absorbed dose), represented in SHEDS by the variable "abstot" on the "allpersons" files. (The abstot variable at the daily or the event level also represents absorbed dose, but it is not the ADD since it is not an average of daily totals.) It is also possible to select a variable that is not time-averaged, such as the maximum one-day dose. 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 resolution 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 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 in each analysis leads to 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 really only meaningful if the difference is not 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.

A standard SHEDS-Multimedia 3 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.

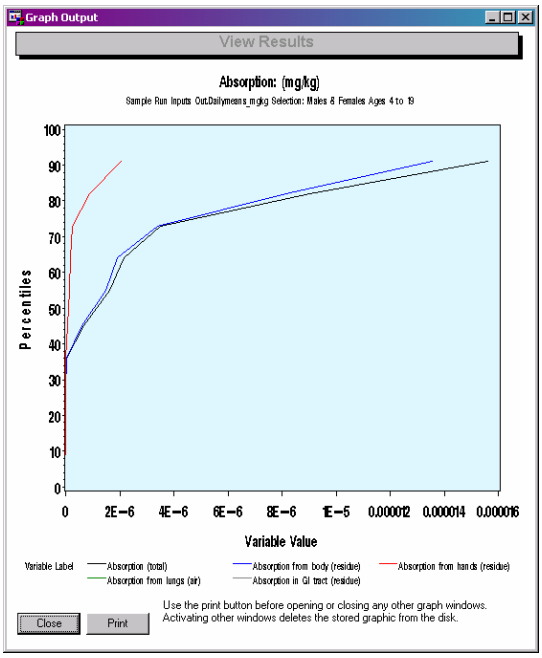


Figure 4-1. Example cumulative distribution function (CDF) plot from a SHEDS variability run.

US EPA ARCHIVE DOCUMENT

Table 4-1. Output files for SHEDS-multimedia version 3 variability run

File	Contents
allpersondays_mgkg	one record for each simulation day, for each person, with exposures and doses in units of (mg/kg/day).
allpersondays_ug	one record for each simulation day, for each person, with exposures and doses in units of micrograms per day.
allvardaily_ug	one record in total, giving the averages across all persons and days for the variables on allpersondays_ug.
daily_mgkg	daily summaries for all output variables, for just the last person simulated, in units of (mg/kg/day).
daily_ug	daily summaries for all output variables, for just the last person simulated, in units of micrograms per day.
dailymeans_mgkg	means of daily summaries for all output variables, averaged across simulation days but not across persons, in units of (mg/kg/day).
dailymeans_ug	means of daily summaries for all output variables, averaged across simulation days but not across persons, in micrograms per day.
diarycounts	number of valid time-activity diaries for each cohort and day type.
dose_mgkg	event-level values for all absorption, blood dose, and elimination variables, for just the last person simulated, in (mg/kg/day).
dose_ug	event-level values for all absorption, blood dose, and elimination variables, for just the last person simulated, in micrograms per day.
expo_mgkg	event-level values for all new exposure, exposure, and absorption variables, for just the last person simulated, in (mg/kg/day).
expo_ug	event-level values for all new exposure, exposure, and absorption, variables, for just the last person simulated, in micrograms per day.
persons	one record for each simulated person, giving time-independent personal data, including age, gender, and input variables that are evaluated once per person.

US EPA ARCHIVE DOCUMENT

5 SENSITIVITY ANALYSES

The sensitivity analysis methods used in SHEDS-Multimedia version 3 are the same ones reviewed or recommended by the December 3-5, 2003 FIFRA Scientific Advisory 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 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)

The SAP reviewed sensitivity analyses involving parametric scaling approaches and a regression approach. The Panel felt that "The regression approach is a more rigorous statistical tool than the two scaling techniques. The multivariate analysis provides a potentially more accurate and useful characterization of the model sensitivities" and recommended that "non-parametric scaling would be a useful and, in some ways, a more consistent way to assess the contributions of different parameters to overall variability (FIFRA SAP, 2004)." SHEDS-Multimedia version 3 uses a nonparametric scaling approach as recommended by the SAP. An additional sensitivity analysis method, Sobol's method, is being considered for SHEDS-Multimedia version 4 (Glen et al., 2007).

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 must 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.

There are two very distinct methods of sensitivity analysis which could be applied to SHEDS. At present, neither method can be operated directly through the user interface, in part because sensitivity analysis is very slow and time-consuming. It may also overtax the computer resources so that some runs fail. Until it is more reliable in practice, sensitivity analysis is considered to be an extra or supplemental feature of SHEDS. The SHEDS User Guide describes how sensitivity analysis can be run in batch mode.

The first of the two sensitivity analysis methods (vary all inputs simultaneously approach) has three variations. It is based on conducting a typical SHEDS variability run, but keeping track for each individual the mean values of the input variables and their corresponding outputs (e.g., ADD). Then the three variations examine the relationship between the output and the various inputs using Pearson correlation, Spearman correlation, and stepwise regression. Each of these is discussed below.

The second type of sensitivity analysis is very different. Instead of allowing all input variables to vary simultaneously, it uses a one-at-a-time percentile scaling approach. Each input sampled from a (non-point) continuous distribution has one of three pre-set values. The "low" value is at the 5th percentile of the input distribution, the "medium" value is at the median or 50th percentile, and the "high" value is at the 95th percentile. A "baseline" run is done with all inputs at their "medium" settings. Since all these inputs are fixed, the output has a reduced variability that reflects other stochastic aspects of the

US EPA ARCHIVE DOCUMENT

modeling. For example, the sampling from probability vectors is still random, as are the demographics, the diary selection, and most yes/no or binary outcome decisions. Then two more runs are done for each input of interest, one with the input at its low setting and the other with the same input at its high setting. All other continuous input variables are fixed at their medium values. Thus, in any run except the baseline, there is exactly one input that is not at its medium value. If N input variables are being analyzed, then (2N+1) runs are needed. From each run, the mean output is determined. The ratio of the mean output for the "high" setting to the mean output for the "low" setting may then be calculated. Also, the ratio of "high" to "medium", or the ratio of "medium" to "low" may be calculated, using the baseline run. Any departures in these ratios from unity must be due either to the effect of changing the input setting or due to stochastic variation remaining in the model.

5.1 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. Hence "y" may be taken to be the logarithm of ADD, as most exposure or dose variables are positively skewed and have logarithms that are closer to being normally distributed.

The 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. 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.

5.2 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. Thus, the values of the output variable are ranked across all the persons, and similarly the values of each input variable are ranked across persons. 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 the "y" and "x" tend to vary in the same direction. 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.

5.3 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 (see Table 5-1). 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.

Table 5-1. Hypothetical stepwise regression and percentile scaling sensitivity analysis results.

			Ratio between 95th and 5th percentiles			
Input Variable	units	Step	Mean	Median	95th	99th
hand mouthing events per hour	events/hr	1	2.22	3.42	1.92	1.84
GI tract absorption rate per day for surface residues	1/day	2	2.20	3.75	1.87	1.70
fraction of body unclothed	[-]	4	2.13	1.43	2.43	2.50
fraction of surface of one hand that enters mouth	[-]	8	1.90	2.83	1.68	1.58
maximum dermal loading for hands	ug/cm^2	9	1.53	1.40	1.54	1.39
removal efficiency during mouthing	[-]	14	1.50	1.80	1.39	1.35
maximum dermal loading for body	ug/cm^2	11	1.45	1.18	1.65	1.72
surface-skin transfer coefficient for body (unclothed)	cm^2/hr	5	1.37	1.34	1.37	1.27
surface-skin transfer coefficient for hand	cm^2/hr	6	1.31	2.00	1.14	1.14
object-mouth contact area	cm^2	20	1.08	1.17	1.06	1.09
soil ingestion rate (outdoor, direct only)	mg/hour	16	1.06	1.25	1.03	0.99
GI tract absorption rate per day for dust or soil	1/day		1.02	1.26	0.98	1.02
dermal absorption rate per day for dust or soil	1/day	13	1.02	1.06	1.01	1.01
dust ingestion rate (indoor, direct only)	mg/hour	21	1.02	1.15	1.00	0.99
object-mouth transfer efficiency	[-]	15	1.01	0.95	1.04	1.00
body-surface fractional contact rate	1/hr		1.01	1.07	1.01	1.00
hand-surface fractional contact rate	1/hr	19	1.01	1.00	1.00	1.00
soil-skin adherence factor	mg/cm^2	18	0.98	1.00	0.99	0.97
residue-skin transfer efficiency	[-]	17	0.97	0.91	0.98	0.99
removal efficiency during hand washing	[-]	10	0.95	0.87	0.96	0.99
removal efficiency during bath/shower	[-]	7	0.77	0.71	0.80	0.86
mean # hand washes/day per person	1/day	12	0.68	0.48	0.77	0.77
elimination rate from the blood	1/day	3	0.52	0.47	0.54	0.53

US EPA ARCHIVE DOCUMENT

5.4 The Percentile Scaling Approach

The second major form of sensitivity analysis is the percentile scaling approach. This differs greatly from the above methods in that the continuous input variables are not allowed to vary randomly, but always use preset values. Each input variable is still assigned a distribution as before, but instead of randomly sampling, the model is told to use one of three values: the "low" being the 5th percentile, the "medium" being the median or 50th percentile, and the "high" being the 95th percentile. The SHEDS model locates the values of these percentiles and assigns them as needed.

The user selects which variables are to be analyzed in a given model run, by setting the variable called "sensitivity" to unity, on the "variables" input file. Variables not to be analyzed are set to sensitivity=0. If N variables are selected, then SHEDS performs a set of (2N+1) model runs, each of the usual number of M persons. The first of these runs, called the baseline run, has all the input variables set to their "medium" values. The M persons do not share the same output because the non-continuous inputs may still differ. For example, the M persons will not necessarily realize the same demographic variables, even though they are drawn from the same population. The application dates may be different, unless the model is run in userdates mode (which is recommended). The activity diaries will likely be different. Finally, the contacts with media are sampled from probability vectors (which are not subject to percentile scaling), and so will be different. Hence, a sample of M persons is needed to represent the exposure distribution resulting from these other effects, even for a fixed set of continuous input variable settings.

For each of the N selected input variables, one model run is performed with this input at its low setting (5th percentile) for all persons, while all other continuous inputs are at their medium settings. Another run is then done with this input at its high setting (95th percentile) for all persons, while all other continuous inputs are at their medium settings. Thus there are two additional runs for each input, apart from the baseline run.

The mean output statistic (for example, ADD) is calculated across the M persons, for each run. 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 (see). In all these ratios, all but one of the inputs are held fixed at its medium value in both runs. Thus, the ratio itself should reflect differences in exposure resulting from the difference in the setting for the one specified input. In practice, the ratios also contain some stochastic variation from the non-continuous SHEDS inputs. The sample size must be large enough that this stochastic variation does not overwhelm the systematic variation from the selected input, or else the latter cannot be quantified. A strong signal that the stochastic variation is too large occurs when the "high/medium" and "low/medium" ratios both exceed unity (or both are less than unity)

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. The principal danger 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. This issue may be resolved in Version 4 by the use of fixed random number seeds.

6 UNCERTAINTY ANALYSES

The uncertainty analysis methods used in SHEDS-Multimedia version 3 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-Multimedia 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 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 (1000 persons for each of the 500 sets of inputs). On typical personal computers, this could take between one week and one month of continuous operation. The actual time can be shortened by running SHEDS simultaneously on several computers and pooling the results.

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. It also does not provide information on the importance of each input individually; this 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 set of inputs. If the inputs had no uncertainty at all, the 500 sets of inputs would all be the same and the 500

US EPA ARCHIVE DOCUMENT

estimates of the 90th percentile would differ only because of random differences in sampling (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 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.

6.1 The Bootstrap Approach for Input Uncertainty Distributions

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.

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.

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. In SHEDS, point distributions are used when all persons are intended to share the same value. For example, the probability of having a lawn may be set to 0.79, which means that each person has a 79% chance of having a lawn (not that the outcome of this test will be the same for all persons). However, this 79% chance may be uncertain, especially if SHEDS is applied to a subpopulation for whom the chance may be different.

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, the bootstrap method could still be used 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 could be used, but this would require 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.

The SHEDS-Multimedia 3 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.

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 sets of inputs. If K<N, then this becomes certain. However, this is not a concern since the other input variables are likely to be different. 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. Even with the unreasonably small value of K=2 for all inputs, there is no danger of

running out of degrees of freedom, as $N << 2^{20}$. In practice, there is virtually no chance that any two of the N sets of inputs will be identical.

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 of individuals at all, since the demographic variables are re-sampled for each of the N input sets. Neither is there a clear correspondence in time, if the application dates are randomly determined. Hence, it is sufficient to summarize each person by some selected statistical parameter before proceeding to the next. For example, the average daily absorbed dose (ADD) may be used for this purpose. In addition, the mean value assigned to each person may be recorded, which permits a combination uncertainty/sensitivity analysis to be performed. See the next section on sensitivity analysis for details on this option.

Without performing sensitivity analysis, there are two commonly used types of graphical analysis for the uncertainty output. One involves displaying the results of three of the N variability distributions, specifically those corresponding to the 5th, 50th, and 95th percentiles of the N cases, as ranked by their medians (Figure 6-1). For example, with N=500 the variability distributions that had the 25th, 250th, and 475th 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 5th, 50th, and 95th) 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.

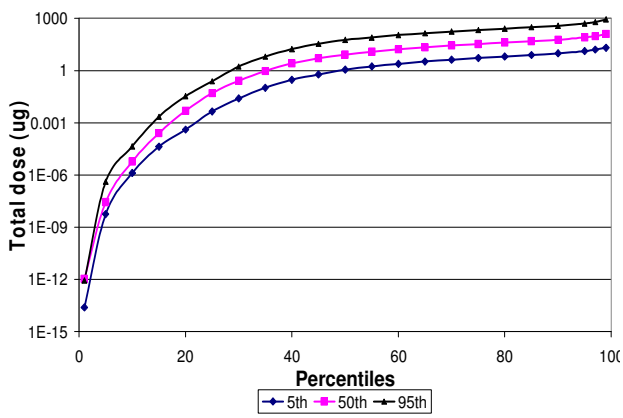


Figure 6-1. Hypothetical uncertainty analysis results: CDFs for low-, medium-, and high-dose populations, as ranked by their medians.

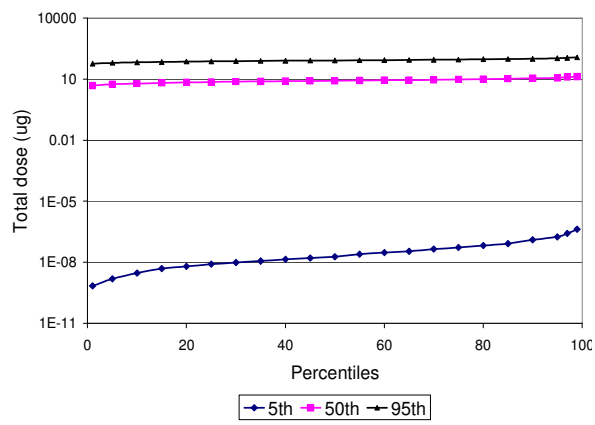


Figure 6-2. Hypothetical uncertainty analysis results: CDFs for the 5th, 50th, and 95th percentiles from all simulated populations.

US EPA ARCHIVE DOCUMENT

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 (see Table 6-1 and Table 6-2). 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.

Table 6-1. Hypothetical Spearman correlation uncertainty analysis results.

Input Variables	R^2
maximum dermal loading for hands	0.0989
maximum dermal loading for body	0.0959
GI tract absorption rate per day for surface residues	0.0921
fraction of surface of one hand that enters mouth	0.0859
fraction of body unclothed	0.0835
Body-surface fractional contact rate	0.0826
removal efficiency during mouthing	0.0823
object-surface concentration ratio	0.0822
probability of having a vegetable garden	0.0822
probability of having a lawn	0.0822
probability of having a dog or cat	0.0822
mass ratio (metabolite/pollutant)	0.0822
object-mouth contact rate	0.0822
removal efficiency during events without water	0.0822
transfer coefficient for object mouthing	0.0822
Hand-surface fractional contact rate	0.0821
removal efficiency during hand washing	0.0821
soil-skin adherence factor	0.0821
dermal absorption rate per day for dust or soil	0.0821
object-mouth transfer efficiency	0.0820
removal efficiency during bath/shower	0.0820
residue-skin transfer efficiency	0.0818
GI tract absorption rate per day for dust or soil	0.0813
elimination rate from the blood	0.0730
dust ingestion rate (indoor, direct only)	0.0487
object-mouth contact area	0.0284
hand mouthing events per hour	0.0246
soil ingestion rate (outdoor, direct only)	0.0076
surface-skin transfer coefficient for hand	-0.0034
surface-skin transfer coefficient for body (unclothed)	-0.0082
mean # hand washes/day per person	-0.0351

US EPA ARCHIVE DOCUMENT

Table 6-2. Hypothetical stepwise regression uncertainty analysis results.

Input Variables	Stepwise Rank	Partial R^2	Model R^2	P value
maximum dermal loading for hands	1	9.78E-03	9.78E-03	0.00
elimination rate from the blood	2	1.15E-02	2.13E-02	0.00
GI tract absorption rate per day for surface residues	3	2.56E-02	4.69E-02	0.00
mean # hand washes/day per person	4	6.65E-03	5.35E-02	0.00
fraction of surface of one hand that enters mouth	5	5.72E-03	5.93E-02	0.00
object-mouth contact rate	6	7.58E-03	6.68E-02	0.00
maximum dermal loading for body	7	5.30E-03	7.21E-02	0.00
fraction of body unclothed	8	2.70E-03	7.48E-02	0.00
hand mouthing events per hour	9	5.09E-04	7.53E-02	0.00
surface-skin transfer coefficient for body (unclothed)	10	1.65E-04	7.55E-02	0.00
removal efficiency during mouthing	11	1.09E-04	7.56E-02	0.00
surface-skin transfer coefficient for hand	12	1.01E-04	7.57E-02	0.00
removal efficiency during bath/shower	13	9.25E-05	7.58E-02	0.00
residue-skin transfer efficiency	14	1.02E-05	7.58E-02	0.01
removal efficiency during hand washing	15	7.59E-06	7.58E-02	0.02
Max mets	16	7.07E-06	7.58E-02	0.02
object-mouth transfer efficiency	17	5.97E-06	7.58E-02	0.04
GI tract absorption rate per day for dust or soil	18	5.92E-06	7.59E-02	0.04

US EPA ARCHIVE DOCUMENT

7 QUALITY ASSURANCE (QA)

SHEDS-Multimedia version 3 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.

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-Multimedia 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. Finally, the complete package was also tested.

Next, individuals in ORD/NERL, OPP, and Alion tested SHEDS-Multimedia, 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.

Finally, SHEDS-Multimedia was released to several interested external beta testers who provided valuable feedback; their comments led to further revisions.

7.2 One-person, one-day 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 one day simulation with selected hand calculations. The equations, inputs used for this, and the resulting outputs are all provided in Appendix G.

A model run of SHEDS-Multimedia covering just one person for one day was made. This individual was a 10-year old girl. An artificial diary day with 30 events was constructed, with enough variety to provide examples of nearly all of the exposure equations. The Post-Application method for chemical concentrations was used, with the concentrations set to integer point values for simplicity. In the exposure code, 'ConcR' is the variable indicating the loading of residues on the contacted surface in $\mu g/cm^2$, 'ConcM' gives the concentration in soil or dust in µg/g, and 'ConcA' gives the air concentration in $\mu g/m^3$. The following lists the values used in this example for each of these variables:

ConcR	=	30	[if residue = 'RYlawn']
	=	10	[if residue = 'RThard']
	=	5	[if residue = 'RTsoft']
	=	5	[if residue = 'RUhard']
	=	2	[if residue = 'RUsoft']
	=	0	[otherwise]

ConcM	=	20	[if matter = 'MTdust']
	=	3	[if matter = 'MUdust']
	=	0	[otherwise]

ConcA	=	2	[if air = 'ATair']
	=	1	[if air = 'AUair']
	=	0	[otherwise]

The data set 'expo_ug' in the output directory was examined in detail, and two tables were constructed to summarize its contents. Table G-1 is 'Exposure_inputs', which lists in alphabetical order the variables input to the exposure module. This table has 30 rows, one for each of the 30 diary events. Each of the 50 columns represents one input variable and are spread over multiple pages. Some of these inputs retain the same value over the course of the day, while others may change hourly or change on every event. Table G-2 provides the exposure equations, and Table G-3 reports the model output.

All of the concentration values (ConcR, ConcM, and ConcA) and all the ingestion values were checked. Sleep events were visually scanned as blocks and no problem was found. Spot checks on other variables were made which included at least one example from every row and every column in Table G-3. All of these checks were successful. Without automating the process (for example, by entering all the data and equations in a spreadsheet), it is not practical to evaluate every item in these tables.

The reader may wish to perform additional spot checks at their own discretion. Instructions for using the equations in Table G-2 in conjunction with the inputs in Table G-1 are provided in Appendix G. These results may then be checked against the outputs reported in Table G-3.

US EPA ARCHIVE DOCUMENT

8 REFERENCES

Cohen Hubal EA, Sheldon LS, Burke JM, McCurdy TR, Berry MR, Rigas ML, Zartarian VG and Freeman NCG. (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.

Cullen AC and Frey HC. (1999). Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs. Plenum Press. New York, NY.

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

Frey HC, 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, Xue J. (2007). Planned Methodologies for Extending SHEDS-Multimedia version 3 (aggregate only) to SHEDS-Multimedia version 4 (aggregate or cumulative). Prepared for August 14-15, 2007 OPP FIFRA Scientific Advisory Panel.

Glen G (2007). SHEDS-Multimedia version 3 SAS code.

Graham SE 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 SW, Yang YC, Chu PL, Sheldon L, Robson M, Needham L, Barr D, Freeman N, Georgopoulos P, Lioy PJ. (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, Glen G, McCurdy T, and Smith L. (2007). Modeling energy expenditure and oxygen consumption in human exposure models: accounting for fatigue and EPOC. Accepted by *Journal of Exposure Science and Environmental Epidemiology.*

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.

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/about/major/nhanes/datalink.htm#NHANESIII.

Smith L and Glen G. (in preparation). Methods for correlating inputs to stochastic models.

Stallings C, Glen G, Smith L, Zartarian VG, Xue J. (2007). SHEDS-Multimedia (Stochastic Human Exposure and Dose Simulation Model for Multimedia, Multipathway Chemicals) version 3 User's

US EPA ARCHIVE DOCUMENT

Guide, prepared for August 14-15, 2007 US EPA/OPP FIFRA Science Advisory Panel (SAP) meeting.

U.S. Census Bureau. (2002a). Census 2000 Summary File 3 – United States.

U.S. Census Bureau. (2002b). Census 2000 Summary File 3 Technical Documentation.

Xue J. (2007). SHEDS-dietary SAS code.

Xue J, McCurdy T, Spengler J, and Özkaynak H. (2004). Intra- and inter-individual activity considerations in human exposure modeling. *J. Expos. Anal. Environ. Epidem.* 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.

Zartarian VG, Özkaynak H, Burke JM, Zufall MJ, Rigas ML, and Furtaw Jr. EJ. (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., J. Xue, H. A. Özkaynak, W. Dang, G. Glen, L. Smith, and C. Stallings. (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.

US EPA ARCHIVE DOCUMENT

APPENDIX A Terminology Used in SHEDS-Multimedia

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	The mass of chemical agent crossing the SHEDS exposure surfaces (skin, GI tract, lungs) into the blood during a given CHAD diary event. This is comprised of dermal absorption (hands or body), GI tract absorption, and lung absorption. The seven variables tracking components of absorption in SHEDS are: absHR, absHM, absBR, absBM, absGR, absGM, and absLA. The sum of these seven absorbed amounts is called "abstot".
Agent	Chemical that contacts the human target and has potential for becoming a dose. SHEDS-Multimedia3 only allows one chemical agent at a time.
Aggregate exposure	Exposures from a single chemical summed over time (but not necessarily over pathways).
Air	The air carrier medium in which the chemical agent may exist, abbreviated "A" in SHEDS code. Only the mass of the chemical is considered. Therefore, the mass of "air" is zero in SHEDS if no chemical is present in the air.
Application	The usage of a product which releases chemical agent. SHEDS allows multiple application types or methods; these may affect various locations and media, and may occur on multiple dates. A given application refers to a specific application type, location, date, and hour. One individual may be subject to multiple applications over the course of the simulation period. Under the SHEDS "modeldates" option, the number, dates and times of applications differ for each simulated individual.
Application type	A specific combination of chemical formulation, application method, and target location. The SHEDS Multimedia 3 interface currently supports 10 application types, discussed in detail.
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

US EPA ARCHIVE DOCUMENT

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.

Bernoulli variable This is a variable with only two possible outcomes. In SHEDS, these variables take the form of a "Yes/No" result. The user inputs a point value for the probability of "Yes". The probability of "No" is automatically determined as (1 - the probability of "Yes").

Body All the skin surface on the body, except the hands; abbreviated "B" in variable names. 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 is considered "unclothed", thereby being available for exposure and loading; this fraction is re-sampled daily.

CHAD The EPA's Consolidated Human Activity Database, described in McCurdy et al. (2000), and available online at http://www.epa.gov/chadnet1 . This is an assemblage containing about 22,000 one-day human activity diaries, and is currently used in several EPA exposure models.

Contact event A CHAD diary event that involves new contact with a (potentially) contaminated contact medium.

Cohort A group of individuals in SHEDS who share the same age and gender. 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.

Contact medium There are three types of contact media in SHEDS: (A) air, (S) surfaces, and (M) matter (dust or soil). Each example that may be expected to have a distinct concentration of the target chemical is considered to be a distinct contact medium. For instance, smooth and textured surfaces may have different retention and transfer properties, so these are different contact media, even in the same room. Furthermore, treated and untreated areas in the house are expected to contain different amounts of chemical, leading to a minimum of four surface contact media in the house (smooth and textured surfaces in each of the treated and untreated areas).

Co-occurrence The potential for multiple applications to affect the same individual in a non-independent fashion. One meaning refers to the timing of multiple applications of the same type; for example, multiple lawn treatments for the same individual.

US EPA ARCHIVE DOCUMENT

In this context, the co-occurrence tends to be exclusionary, as multiple treatments tend not to occur too near each other in time. The second meaning of "co-occurrence" refers to the possibility of more than one application type happening on the same (or nearly the same) date; for example, both the lawn and garden may be treated. Here the co-occurrence often enhances the likelihood of multiple applications occurring close together in time.

Concentration Time Series A time series with time on the x-axis and contact media concentration on the y-axis.

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.

Decay/dispersion One of the built-in source-to-concentration modules in SHEDS. The treated area undergoes first-order exponential decay in concentration, while nearby untreated areas may acquire chemical via dispersion from the treated area.

Dermal SHEDS separates the skin into two regions: hands and body.

Diary event The time basis on which all exposure and dose variables are calculated in SHEDS. Diary events have variable duration, from 1 to 60 minutes. Diary events never cross the top of a clock hour. The user may lower this upper bound for duration on the "Specify Simulation Information" screen of the model interface. The number of diary events in the simulation period generally differs between individuals.

Distribution One of two ways for specifying the form of stochastic or randomly-determined inputs (the other being *probability vector*). Distributions are used for continuous variables in SHEDS, which may be sampled from the beta, gamma, exponential, lognormal, normal, point, triangle, uniform, and Weibull statistical distributions.

Dose The mass of chemical absorbed (absorbed dose) or metabolite in the blood (blood dose) after the chemical agent has crossed the exposure surfaces. The pharmacokinetic model built into SHEDS assumes that upon absorption, all of the mass is immediately converted to a single metabolite and stored in the blood. The input variable "metab_ratio" gives the mass ratio of the metabolite to the

US EPA ARCHIVE DOCUMENT

parent compound (the chemical agent). Setting metab_ratio=1 means that the dose measures the parent chemical. Each event, the seven blood doses are each increased by corresponding "abs" term (multiplied by metab_ratio), and reduced by the corresponding "elm" term. Blood dose is measured in micrograms of metabolite (µg). When aggregating to longer time periods, doses must be averaged, not summed.

Elimination The amount of metabolite (or parent chemical if metab_ratio=1) eliminated from the blood during the current CHAD diary event. The seven elimination variables are: elmHR, elmHM, elmBR, elmBM, elmGR, elmGM, elmLA. Elimination is considered in micrograms of metabolite (µg) per diary event, and modeled as a first-order exponential decay process. It is intended to represent filtration by the kidneys and/or chemical transformation in the liver. It does not directly correspond to urinary excretion.

Exposure Contact between a chemical agent and human target at the exposure surfaces (skin, GI tract, lungs). SHEDS distinguishes exposure in three ways: by the affected body part, by the form in which the chemical is present (dust/soil, residue, or air), and by whether the exposure is "new" or "running." New exposure only counts that which is transferred to the body part during the current diary event. "Running exposure" is the total mass of chemical currently on the body part, resulting from all prior additions and subtractions.

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 ingestion pathways it is the gastrointestinal (GI) tract. For inhalation exposure it is the lungs.

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.

GI tract The gastrointestinal tract, abbreviated "G" in SHEDS variable names. Exposure from food, water, hand mouthing, and object mouthing can lead to GI tract exposure. Currently, food and water are handled by a separate SHEDS dietary module, and are not integrated into SHEDS-Multimedia version 3.

Hands The skin area below the wrists on both hands, abbreviated "H" in 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 drawn from a user-specified input distribution, which is expressed as the mean number of hand washes per day. This probability-based method means that the particular times and the frequency of hand-washing events vary

US EPA ARCHIVE DOCUMENT

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 A quantity used in the determination of exposure or dose that is randomly sampled according to user-specified rules. It is stochastically sampled either from a distribution (for continuous variables) or from a probability vector (for discrete variables). See also the *sampling frequency*.

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.

Lungs The lung exposure surface in SHEDS which may receive exposure to the chemical agent via inhalation, abbreviated "L" in SHEDS variable names.

Matter A generic name for soil or dust, abbreviated "M" in variable names. In SHEDS, "soil" is any non-residue matter that adheres to the skin and is encountered outdoors. "Dust" is similar, except that it is encountered indoors. Thus, dirt on the shoes is "soil" when it is picked up outdoors, but it is "dust" when deposited on the floor of the house. SHEDS currently combines soil and dust into a single total. Like "air" and "residue," measures of "matter" in SHEDS refer only to the mass of the target chemical.

Micro A short form for "microenvironment", which in SHEDS indicates the individual's immediate surroundings for the duration of the current CHAD diary event (note that this term is defined differently in non-SHEDS contexts). SHEDS uses five micros: in_home, out_home, in_other, out_other, and in_vehicle. The micro is determined from the CHAD location code. Some micros may be subdivided stochastically, such as the division of "in_home" into "treated room", "untreated room", and "far from source" (see sub-micro).

New exposure The amount of chemical agent transferred from a medium by a particular pathway to the person target, on a given diary event, abbreviated "exp". SHEDS currently tracks seven distinct exposure variables: expHR, expHM, expBR, expBM, expGR, expGM, expLA. In SHEDS the "exp" variables are considered in micrograms of target chemical (per diary event), usually abbreviated "µg."

Non-dietary ingestion Ingestion of chemical through hand or object mouthing.

US EPA ARCHIVE DOCUMENT

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 Following (but not during) a residential chemical application. Also refers to one of the built-in methods for source-to-concentration modeling in SHEDS.

Probability vector One of two alternatives for specifying the form of a stochastic input (the other being a *distribution*). It is used for variables that have more than two discrete ~~set of~~ outcomes. Each outcome is assigned a probability. These should total to one, as each time the vector is sampled, one of the outcomes must occur.

Residue One of the phases in which the target chemical may be present, abbreviated "R". Generally, it refers to chemical residue that may be encountered when a surface is contacted.

Running exposure The current mass of chemical on/in the given body part. This includes both new exposure and the amount carried over from prior events. The ones currently tracked in SHEDS are: ldgHR, ldgHM, ldgBR, ldgBM, ldgGR, ldgGM, ldgLA. All loadings are in micrograms (µg) of chemical. Care must be taken not to confuse "exp" and "ldg" variables.

Sampling frequency 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, hourly, or every event. Such variables change values over time for a single person, as well as differing between persons.

Simulation period 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.

US EPA ARCHIVE DOCUMENT

Stochastic An adjective meaning "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.

Sub-micro A sub-microenvironment is a division of a microenvironment (see "micro") into regions of differing concentration, usually because one part is "treated" while the rest is "untreated". The "in-home" micro can optionally be divided into a third sub-micro called "neither", which can indicate other houses than the one being treated. While the micro is determined deterministically by the codes on the CHAD diary, the sub-micro is determined probabilistically using distributions input into SHEDS.

Target In SHEDS, a simulated person that receives an exposure.

Treated room 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.

Untreated room An area of a simulated individual's residence which is not directly treated, but which may become contaminated by diffusion/dispersion from a treated area.

US EPA ARCHIVE DOCUMENT

APPENDIX B Selected SAS Variable Nomenclature

abs	new dose
absr_dm	dermal absorption rate per day for dust or soil
absr_gm	GI tract absorption rate per day for dust or soil
absr_gr	GI tract absorption rate per day for surface residues
absf_lung	absorption fraction for lungs
adherence	soil-skin adherence factor
bioavm	bioavailability fraction for dust/soil
bioavr	bioavailability fraction for surface residues
BM	dust or soil on body
BR	residues from surfaces on body
contactb	body-surface fractional contact rate
contacth	hand-surface fractional contact rate
dermaxb	maximum dermal loading for body
dermaxh	maximum dermal loading for hands
dos	running dose
dosBR	dose from body residues
dosBM	dose from body soil/dust
dosGR	dose from GI tract residues
dosGM	dose from GI tract soil/dust
dosHR	dose from hand residues
dosHM	dose from hand soil/dust
dosLA	dose from inhalation
elimr_blood	elimination rate from the blood
elm	elimination
exp	new exposure
f_uncloth	fraction of body unclothed
GM	dust or soil in GI tract
GR	residues from surfaces in GI tract
has_garden_p	probability of having a vegetable garden
has_lawn_p	probability of having a lawn
has_pet_p	probability of having a dog or cat
HM	dust or soil on hands
hm_fraction	fraction of surface of one hand that enters mouth
hm_freq	frequency of hand mouthing events per hour
HR	residues from surfaces on hands
InCarRm	Indoor carpet and room (aerosol)
InCCaer	Indoor crack and crevice (aerosol)
InCCliq	Indoor crack and crevice (liquid)
InFIK	Indoor flying insect killer (aerosol)
InFog	Indoor fogger (broadcast)
ingestion_indoor	dust ingestion rate (indoor, direct only)
ingestion_outdoor	soil ingestion rate (outdoor, direct only)
in_home	inside a residence
in_oth	inside a building other than a residence

US EPA ARCHIVE DOCUMENT

in_veh	in a vehicle
LA	inhalation pathway
LawnGra	Lawn (granular - push spreader)
LawnLiq	Lawn (liquid - handwand)
ldg	running exposure
metab_ratio	mass ratio (metabolite/pollutant)
object_ratio	object-surface concentration ratio
om_area	object-mouth contact area
om_freq	object-mouth contact rate
om_transfer	object-mouth transfer efficiency
out_home	outside around home
out_oth	outside away from home
Pet	residues on pet fur
PetLiq	Pet treatment (liquid)
PetSpot	Pet treatment (spot-on)
probN_in_home	probability for being too far from a treated area for any chemical to be present
probT_in_home	probability for being in a treated room
probU_in_home	probability for being in an untreated room
remv_bath	removal efficiency during bath/shower
remv_dry	removal efficiency during events without water
remv_mouth	removal efficiency during mouthing
remv_wash	removal efficiency during hand washing
tc_b	surface-skin transfer coefficient for body (unclothed)
tc_h	surface-skin transfer coefficient for hand
tc_m	transfer coefficient for object-mouthing
Tair	air in treated room
Tdust	dust in treated room
Thard	smooth (hard) surface residues in treated room
transfer_dermal	residue-skin transfer efficiency
Tsoft	textured (soft) surface residues in treated room
Uair	air in untreated room
Udust	dust in untreated room
Uhard	smooth surface residues in untreated room
Usoft	textured surface residues in untreated room
user_pr	the likelihood of actively using the application type, given that the appropriate requirement has been met (i. e., having a lawn, garden, or pet)
VegDust	Vegetable garden (dust, powder)
washprob	mean # hand washes/day per person
Yair	air outside residence
Ylawn	surface residues on lawn
Ysoil	soil in treated outdoor locations
Yveg	surface residues in vegetable garden

US EPA ARCHIVE DOCUMENT

APPENDIX C The SHEDS 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. Figure C-1 illustrates the general structure of SHEDS-Multimedia version 3 and how the modules are linked. The SHEDS code is written in SAS. All of the code is together is in a single file, but internally this code is divided into several modules. Each step outlined above consists of one or more modules. Step 1 is handled by the "Persons" module, which assigns the variables that are taken to be constant (time-independent) for each 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.

Step 2 is handled by the "DiarySelection" module. This involves choosing a set of one-day activity diaries from the CHAD database and concatenating them according to the sequence of day-types in the simulation period (with this sequence being determined by the "Calendar" module). The result is 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.

Step 3 has several variations, depending on the method specified for determining the concentrations. There are three main options. The simplest (conceptually) is to let the user directly supply the concentration time series for each contact medium. The other two options first require the determination of chemical usage patterns for the given household. The "UserDates" option requires the user to specify the calendar dates when chemicals are applied. These dates are taken to be the same for all the persons in the model run; however, the amounts of chemical used may vary. This mode of operation is useful for certain types of questions, such as determining the variation in exposure resulting from a single usage event at the start of the simulation. The final option is called "ModelDates", in which the model randomly selects chemical usage (application) dates for each individual. The user supplies the distributions for these usage patterns by specifying, for example, the likelihood of applying a pesticide to the lawn during each month of the year. The modules that handle these three options are called "AppTimeSeries", "AppUserDates", and "AppModelDates", respectively. In all three cases, another module called "Concentrations" actually builds and stores a concentration time series for each potential contact medium. The concentration time series are usually hour-by-hour, over the simulation period. There are typically 10-20 such time series, each containing 8760 hourly values (for a 365-day simulation). All of these are recalculated for each individual simulated.

US EPA ARCHIVE DOCUMENT

US EPA ARCHIVE DOCUMENT

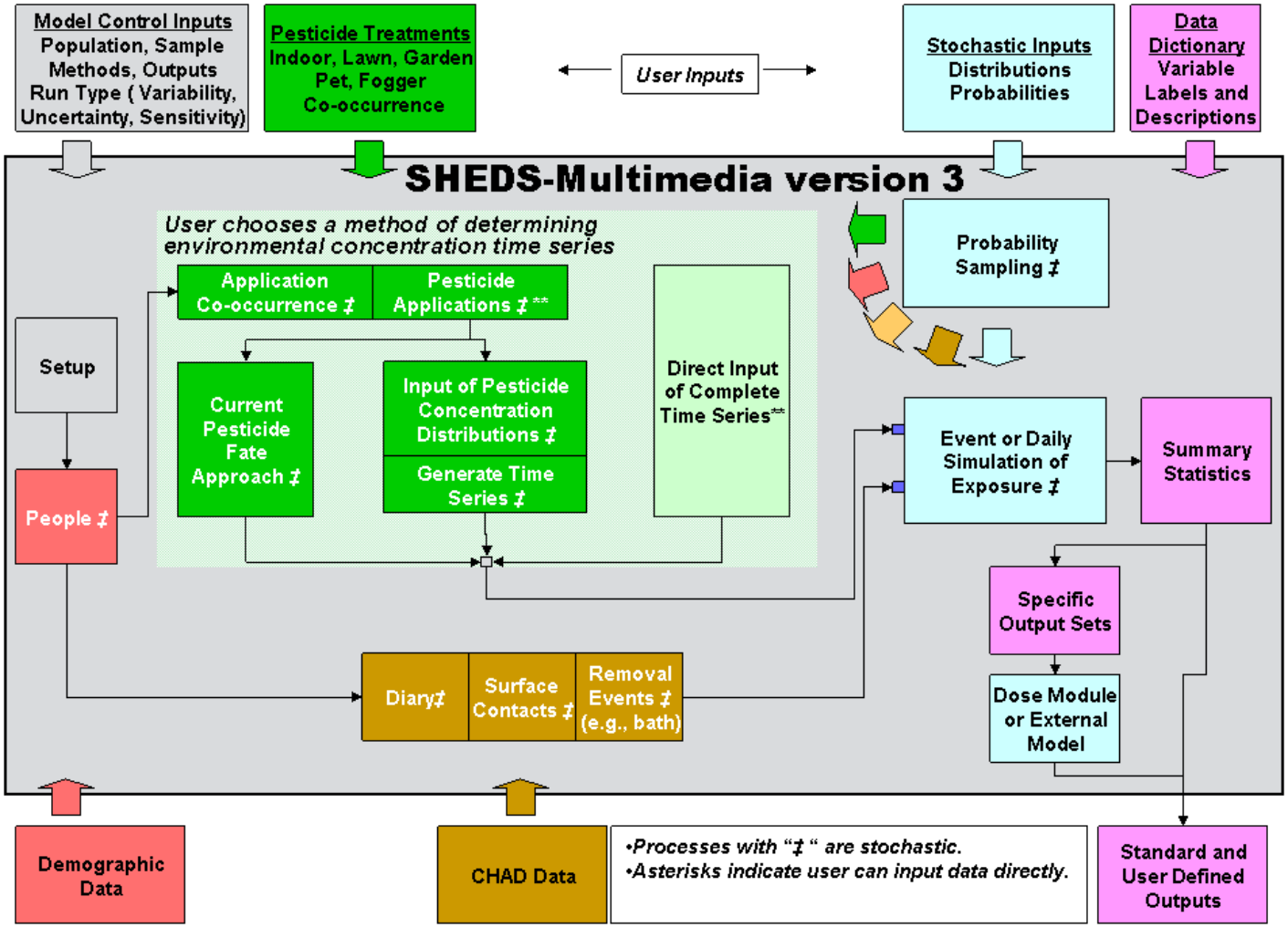


Figure C-1. Conceptual overview of SHEDS-Multimedia version 3 module linkages.

Step 4 is the function of the "DiaryDetailing" module. The activity diaries contain information on place and time and a general description of the main activity, but are short on details. Thus, the diaries may indicate that a child is playing in a given room for 30 minutes, but SHEDS must determine which of the (contaminated) surfaces are contacted during that event. For modeling purposes, all areas in houses are divided into three categories: treated rooms, untreated rooms, and far from sources. Treated rooms are those in which chemical has been (or will be) used. Note that a person may enter the "treated room" prior to the chemical usage, at which time the concentrations are very low or zero. Untreated rooms are those in which the target chemicals are never used directly, but into which some chemical may flow from nearby treated rooms. The third possibility (far from source) means that no chemical is present, and no new contacts can occur. This category may or may not be used, at the user's preference.

It could apply, for example, to time spent in another person's house, which is in the "indoor-residential" category, but has no potential for contact. The DiaryDetailing module determines which surface medium (if any), which air medium (if any), and which dust or soil medium (if any) are contacted on each diary event. At most, only one medium of each type (surface, air, dust/soil) may be contacted on a single diary event. Note that only the media with nonzero pollutant concentrations are normally considered, so "no air contact" means that the person was in uncontaminated air.

An important principle in SHEDS is that steps 2, 3 and 4 are entirely separate. Thus, the chemical concentrations cannot be affected by the choice of activity diaries, or by the selected contact events. Conversely, contact events are not affected by chemical concentrations. There is one partial exception to this rule; after each chemical usage there may be a "re-entry prohibition period", during which the individual cannot contact the pollutant. It is assumed that people are either avoiding or are not permitted to be in the treated area during this period. Since modules 3 and 4 do not directly communicate, the re-entry prohibition is handled in step 5, below.

Step 5 is the "Exposure" module, which is the conceptual core of the SHEDS code. Here, the detailed activity diary with all its events is merged with the concentration time series for the relevant contact media. A "contact event" in SHEDS is a diary event in which new contact with one or more affected media occurs. Contact events result in new exposure. Contact events with air result in lung or inhalation exposure. Other contact events result in hand or body exposure. Here, "hand" means dermal loading on either or both hands. "Body" refers to dermal loading anywhere except on the hands. Gastrointestinal (GI) tract exposure may result from direct ingestion of soil or transfer of chemical from hand-to-mouth activity.

SHEDS distinguishes and tracks both "new exposure," which is the mass of chemical added during the current contact (diary) event, and "running exposure," which reflects additions and removals of chemical from exposure surfaces from the current and prior events. The latter incorporates the loading that is "carried over" from prior events. In SHEDS, the chemical resides on the skin until it is dermally absorbed, transferred to the mouth, washed off, or removed some other way (for example, by "brush-off" or by exceeding the maximum dermal loading). One consequence of this methodology is that SHEDS can determine the benefit of (say) washing the hands after coming in from an outdoor play event, by reducing the dermal loading before it enters the body. However, in SHEDS, when the word "exposure" is used to mean the running exposure mass, note that this exposure will generally not be zero, even when the person is a long way from the chemical sources and no contact is occurring. In this case, the exposure is entirely due to carryover from prior events.

Step 6 is handled by the "Dose" module. This is an optional step in SHEDS. The Dose module is a simple calculation of the current blood dose, taking into account the prior dose, the additions to the dose from adsorption, and the reductions due to elimination (modeled as a first-order exponential decay). If the user prefers, the exposure time series from SHEDS could be used as input to another dose model.

US EPA ARCHIVE DOCUMENT

The final step in the modeling of a person is handled by the "PersonSummary" module. This calculates summary statistics from the various exposure and dose time series. Due to the amount of data produced when simulating even just one person, it is impractical to retain all of the information generated during a model run. This can amount to 100-200 variables for each diary event, or several million values per person. In general, the detailed exposure and dose time series are not retained, but are overwritten by the results for the next person simulated. Therefore, the pertinent information must be extracted before proceeding to the next person.

The SHEDS code also contains other modules for various purposes, such as controlling the overall program flow or for performing certain specialized functions, such as drawing random samples from distributions. The code makes extensive use of the SAS macro language.

C.1 List of SAS Code Modules and Their Calling Sequence

The flow of the SHEDS-Multimedia program code can be understood by examining the body of the top-level module, which contains the following:

```
%macro multimedia3(maindir,runname);
   %Start;
   %Calendars(calendar,&beg_y,&beg_m,&beg_d,&end_y,&end_m,&end_d,0);
   %if (&diaryproc=yes) %then %do; %DiaryPreProcess; %end;
   %CreateAppMedia;
   %Parm_setup;
   %do loop=1 %to &lastloop;
      %StartLoop;
      %Parm_eq;
      %Persons;
      %DiaryGroups;
      %App&AppOption;
      %do person=1 %to &numpersons;
         %Concentrations;
         %DiarySelection;
         %DiaryDetailing;
         %Exposure;
         %Dose;
         %PersonSummary;
      %end;
      %LoopSummary;
   %end;
   %FinalSummary;
%mend multimedia3;
```

Each time the Run button in the interface is pressed, or each time a batch job is submitted, the Multimedia3 macro is called once. Macros in SAS resemble subroutines in many other languages, except perhaps in the rules for passing and sharing variables. Macros are invoked by a percent sign (%) followed by the name of the macro. Only one of the macros listed here (%Calendars) has an explicit argument list, as the others operate on a list of global variables defined in %Start.

The first four (or five, if diary preprocessing is invoked) macros are executed only once per run. Then the first or outer DO loop is started, which processes all the persons in one loop of the uncertainty or sensitivity iterations. For a single-stage (variability-only) run, this outer loop is only processed once. The next five macros from %Startloop through %App&Appoption are executed once for each sample population. In this set, the %Persons macro defines the characteristics of all the individuals or persons in

the sample population. The %App&Appoption macro determines the chemical usage patterns and specific application dates for all these persons.

The inner DO loop is executed once for each person in the sample population. The %Concentrations macro determines the time series of chemical concentration in each of the contact media. These concentrations differ from person to person due to differences in chemical usage as well as differences in other personal characteristics such as house size. %DiarySelection and %DiaryDetailing assemble a longitudinal activity pattern from one-day CHAD diaries, and add details not present on the CHAD diaries, like specific contacts with various media. The %Exposure macro brings together the environmental concentrations with the personal activity patterns and contact sequences. The %Dose macro invokes the internal PBPK model. Finally, summary statistics for each person are calculated and saved in the %PersonSummary macro. At the end of both the inner and outer DO loops, additional summary statistics are calculated.

Very few of the above macros call other macros. There are 11 very short macros that operate essentially as programmer-defined functions. The complete list of these "utility macros" is the following:

%len	Counts the number of items in a list
%checkinputfiles	Returns a message if the agegroups file is missing
%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
%reentry	Converts re-entry time category to a duration

Apart from calls to these utility macros, the only other nested macro calls in SHEDS-Multimedia are the following:

- If the user has requested that the randomly-drawn samples from two or more input variables be correlated, then %Parm_setup internally calls %CorrelSetup.
- The call to %App&AppOption may invoke any one of three macros, depending on the setting for &AppOption: namely, it could be %AppModelDates, %AppUserDates, or %AppTimeSeries. The macro %AppModelDates calls %Calendars, %SetDailyProbs, and %DetailOneApp. The other two do not have any internal macro calls except for the utility macros.
- The %Concentrations macro calls %Calendars if the TimeSeries option is used.
- The %DiaryDetailing macro calls %Mets, which assigns inhalation rates.

The above four cases are the only examples of nested macro calls (other than calls to the utility macros) in SHEDS-Multimedia version 3. Other than the utility macros, the only macro that is called from more than one place in the code is the %Calendars macro.

US EPA ARCHIVE DOCUMENT

APPENDIX D Summary of the CHAD Database

The Consolidated Human Activity Database (CHAD) was created by compiling 12 studies into a single human activity pattern database resulting in 22,968 diary days (McCurdy et. al., 2000). 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, or with time sleeping were dropped. The SHEDS installation package contains 21,114 diaries, but 542 of these have age=0 and cannot currently be used by the model, so effectively the database contains 20,572 diaries. The following table lists the 12 studies, their dates, age ranges, and the number of diaries available for SHEDS.

Table D-1. CHAD Studies used in SHEDS.

Study	Dates	Ages	Diary Count
Baltimore	1997-1998	77-88	42
California: Adults	1987-1988	18-94	1,534
California: Youth	1987-1988	12-17	180
California: Children	1989-1990	1-11	1,086
Cincinnati	1985	1-86	2,282
Denver	1982	18-70	584
Los Angeles: Elementary School	1989	10-12	37
Los Angeles: High School	1990	13-17	38
National: NHAPS-A	1992-1994	1-93	4,570
National: NHAPS-B	1992-1994	1-93	4,521
University of Michigan: Children	1997	1-13	4,956
Valdez, Alaska	1990-1991	11-68	201
Washington, DC	1982-1983	18-71	541

SHEDS groups activity diaries into 10 age categories. The following table provides the age-gender breakout of CHAD diaries available to SHEDS.

Table D-2. Diary Counts for SHEDS Age-Gender Cohorts.

Age / Gender	1	2	3-5	6-10	11-15	16-20	21-29	30-49	50-69	70+
Female	359	353	947	1,506	900	474	1,148	2,725	1,719	731
Male	344	368	1,089	1,627	962	459	1,060	2,162	1,250	389

US EPA ARCHIVE DOCUMENT

The following table 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

US EPA ARCHIVE DOCUMENT

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

US EPA ARCHIVE DOCUMENT

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

US EPA ARCHIVE DOCUMENT

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

US EPA ARCHIVE DOCUMENT

Table D-5 provides an exhaustive list of CHAD location codes and their descriptions.

Table D-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

US EPA ARCHIVE DOCUMENT

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 locat
33500	At School
33600	At Restaurant
33700	At Church
33800	At Hotel /Motel
33900	At Dry cleaners

US EPA ARCHIVE DOCUMENT

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

US EPA ARCHIVE DOCUMENT

APPENDIX E Probability Density Functions

The explicit probability density functions (pdf's) 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 $\sigma^2 < \mu (1-\mu) \leq 1/4$.

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

$$p(x) = \text{Exp}[-(x-v1)/(v2-v1)] / (v2-v1), \quad \text{for } x>v1$$

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 v1. 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 $\sigma = v2\ \text{Sqrt}(v1)$. The PDF of the SHEDS gamma is

Equation E-3

US EPA ARCHIVE DOCUMENT

$p(x) = v2^{-v1}\ x^{v1-1}\ \mathrm{Exp}(-x/v2) / \Gamma(v1)$, for $x>0$

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

$p(x) = \mathrm{Exp}[(-1/2)\ (\mathrm{Log}[x/v1] / \mathrm{Log}[v2])^2\ ] / (x\ \mathrm{Sqrt}[2\ \pi\ ]\ \mathrm{Log}[v2]\ )$, for $x>0$

If GM and GSD are given, then the lognormal has arithmetic mean and standard deviation

Equation E-5

$\mu = \mathrm{GM}\ \mathrm{Exp}[(1/2)\ (\mathrm{Log}(\mathrm{GSD}))^2\ ]$,

$\sigma = \mathrm{GM}\ \mathrm{Sqrt}[\mathrm{Exp}(\ [\mathrm{Log}(\mathrm{GSD})]^2)\ (\mathrm{Exp}(\ [\mathrm{Log}(\mathrm{GSD})]^2)-1)]$.

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

$\mathrm{GM} = \mu / \mathrm{Sqrt}(1 + \sigma^2 / \mu^2)$,

$\mathrm{GSD} = \mathrm{Exp}(\mathrm{Sqrt}(\mathrm{Log}(1 + \sigma^2/\mu^2\ )))$.

If instead, one has the mean μ_{log} and standard deviation σ_{log} of log(x), then use

Equation E-7

$\mathrm{GM} = \mathrm{Exp}(\mu_{log})$,

$\mathrm{GSD} = \mathrm{Exp}(\sigma_{log})$.

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

$p(x) = \mathrm{Exp}[-(x-v1)^2/(2\ v2^2)] / (\mathrm{Sqrt}[2\ \pi]\ v2)$

E.6 Point

US EPA ARCHIVE DOCUMENT

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 $\mu = (v1+v2+v3)/3$, which coincides with the peak only when v2 is midway between v1 and v3. The standard deviation is $\sigma = \mathrm{Sqrt}[(v1^2+v2^2+v3^2-v1v2-v1v3-v2v3)/18]$. 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

$$p(x) = 2\,(x-v1) / [(v2-v1)\,(v3-v1)], \quad \text{for } v1 \le x \le v2$$

$$= 2\,(v3-x) / [(v3-v2)\,(v3-v1)], \quad \text{for } v2 \le x \le v3$$

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 $\sigma = (v2-v1) / \mathrm{Sqrt}(12)$. 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 $\sigma = v2\ \mathrm{Sqrt}[\Gamma(1+2/v1) - (\Gamma(1+1/v1))^2]$. Here 'Γ' 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

$$p(x) = v1\ v2^{-v1}\ x^{v1-1}\ \mathrm{Exp}\,[-(x/v2)^{v1}], \quad \text{for } x>0.$$

E.10 Discrete Probability Density Functions

Bernoulli 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 Bernoulli variables by having the user specify only the probability of a "yes". The probability of "no" is implied by 1 - "yes".

Multinomial variables allow more than two possible outcomes. The usual statistical notation would be (p_1, p_2, …, 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.

US EPA ARCHIVE DOCUMENT

APPENDIX F Description of SHEDS Multimedia Input Files

F.1 List of Input Files

The following are the eight key input data files that the user works with to run the SHEDS-Multimedia version 3 model:

variables:	contains names, units, and descriptions of personal input variables, exposure factors, dose factors, and the re-sampling frequencies for these variables
distributions:	contains one or more probability distributions for each variable listed in the variables file, and the conditions indicating when each of these distributions should be used
correlations:	contains pairs of stochastic input variables and the desired Spearman correlation between them
contactmedia:	contains information on all the contact media (surfaces, air, or matter) which may expose a person to the target chemical
modeldates:	the list of application types and the rules for having the model randomly determine the dates when these applictions occur
userdates:	the list of application types and the specific dates on which these occur (same dates for all persons)
appmedia:	a complete list of all combinations of application types and affected contact media that can be included in a model run
diarydetails:	provides distributions for the contact probabilities for the various contact media.

In addition, the following files are used internally by the model. These are applicable for a wide range of 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.
diarycohorts:	This file is used to group the activity diaries into pools.
diaryevents:	This is the large file containing all the diary events from all persons in the CHAD database.
diarylevels:	This file contains the probability matrix for selecting CHAD diaries with similar amounts of outdoor time, based on the overall tendency assigned to each individual.
diarylocs:	This file contains the mapping from CHAD location codes to the microenvironments used in SHEDS.
diarymets:	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.
diaryQArules:	This file contains a list of logical rules for accepting a CHAD diary for use in SHEDS-Multimedia 3.
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.
pop2000:	This file contains a 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 titled "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 following SAS variables:

Table F-1. Column definitions for the variables file.

column name	format	purpose
variable	$24	name of input variable, as appearing in the code
label	$90	description of the input variable
module	$8	name of SHEDS module where the input is used
priority	num	order in which the records are to be processed
freq	$8	the resampling frequency
units	$16	the units for this input variable

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 default version of the variables file contains 40 records; that is, information on the following 40 input variables:

Table F-2. Definitions for variables from the distributions and variables files.

Variable	Description	Frequency	Units
absf_lung	absorption fraction for lungs	person	[-]
absr_dm	dermal absorption rate per day for dust or soil	person	1/day
absr_dr	dermal absorption rate per day for surface residues	person	1/day
absr_gm	GI tract absorption rate per day for dust or soil	person	1/day
absr_gr	GI tract absorption rate per day for surface residues	person	1/day
adherence	soil-skin adherence factor	person	mg/cm2
bathdays	maximum number of days between baths	person	days
bioavm	bioavailability fraction for dust/soil	person	[-]
bioavr	bioavailability fraction for surface residues	person	[-]
contactb	Body-surface fractional contact rate	hour	1/hr
contacth	Hand-surface fractional contact rate	hour	1/hr
dermaxb	maximum dermal loading for body	person	ug/cm2
dermaxh	maximum dermal loading for hands	person	ug/cm2
elimr_blood	elimination rate from the blood	person	1/day
f_uncloth	fraction of body unclothed	day	[-]
has_garden_p	probability of having a vegetable garden	person	[-]
has_lawn_p	probability of having a lawn	person	[-]
has_pet_p	probability of having a dog or cat	person	[-]
hm_fraction	fraction of surface of one hand that enters mouth	person	[-]
hm_freq	hand mouthing events per hour	hour	events/hr
ingestion_indoor	dust ingestion rate (indoor, direct only)	person	mg/hour
ingestion_outdoor	soil ingestion rate (outdoor, direct only)	person	mg/hour
metab_ratio	mass ratio (metabolite/pollutant)	person	[-]
object_ratio	object-surface concentration ratio	person	[-]
om_area	object-mouth contact area	person	cm2
om_freq	object-mouth contact rate	hour	events/hr

US EPA ARCHIVE DOCUMENT

Table F-2. Definitions for Variables from the Distributions and Variables Files (continued).

Variable	Description	Frequency	Units
om_transfer	object-mouth transfer efficiency	hour	[-]
reentry_indoor	probability vector for reentry time indoors	day	[-]
reentry_outdoor	probability vector for reentry time outdoors	day	[-]
remv_bath	removal efficiency during bath/shower	hour	[-]
remv_dry	removal efficiency during events without water	person	1/hr
remv_mouth	removal efficiency during mouthing	hour	[-]
remv_wash	removal efficiency during hand washing	hour	[-]
tc_b	surface-skin transfer coefficient for body (unclothed)	hour	cm2/hr
tc_h	surface-skin transfer coefficient for hand	hour	cm2/hr
tc_m	transfer coefficient for object mouthing	hour	cm2/hr
timeofday_indoor	probability vector for hour of indoor application	day	[-]
timeofday_outdoor	probability vector for hour of outdoor application	day	[-]
transfer_dermal	residue-skin transfer efficiency	hour	[-]
washprob	mean # hand washes/day per person	person	1/day

F.2.2 Distributions file

The "distributions" file contains the distributions from which the input variables are to be sampled, for all the ones listed on the "variables" file. The "distributions" file contains the following SAS variables:

Table F-3. Column definitions for the distributions file.

Column Name	Format	Purpose
variable	$24	name of input variable, as appearing in the code
condition	$200	when this distribution is utilized for this input variable
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)
minval	num	the minimum allowed value for this variable
maxval	num	the maximum allowed value for this variable
vprobs	$200	a comma-separated list of probabilities
function	$200	used for probability-vector output and special functions
n	num	number of points in uncertainty vectors
v1_un	$3000	comma-separated list of v1 values in uncertainty vector
v2_un	$3000	comma-separated list of v2 values in uncertainty vector
v3_un	$3000	comma-separated list of v3 values in uncertainty vector

For a variability run, the last four columns are not relevant. If the condition is left blank, the same distribution is always used whenever this variable is evaluated. If more than one distribution is intended for a given input variable, then the condition variable must 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.

The "form" variable indicates the type of distribution. It is not case-sensitive, and may be set to the following:

Form	**Purpose**
point	a single point value
uniform	a random value from a uniform distribution
normal	a random value from a normal distribution
lognormal	a random value from a lognormal distribution
triangle	a random value from a triangle distribution
exponential	a random value from an exponential distribution
beta	a random value from a beta distribution
gamma	a random value from a gamma distribution
weibull	a random value from a Weibull distribution
probvector	a random value from a probability vector
function	a value defined by the specified function

The minval is the lower truncation point for the distribution. If a value is generated below this, it is replaced by minval. Similarly, maxval is the upper truncation point, the largest value that may be returned. If the minval or maxval are not defined, then no truncation occurs at that end.

Vprobs lists the probabilities for each outcome in a probability vector. The code divides each element in the list by the sum before using them, so in practice the vector is normalized to unit probability. The interface insists on having the sum equal to unity, so the model is actually using the same values as the user input. The "function" variable has two uses. For probability vectors it gives the return value when each cell is selected. Usually, this is just the cell number itself, but it could be set to a character string, for example. This would probably necessitate code modifications to make some subsequent action depend on such strings, so most users have no use for this feature. The second purpose of "function" is to allow the user to specify distributions other than the standard types. For example, if "variable" was set to "absf_lung", and "form" was set to "function", and the "function" variable were set to "Rand("CHISQUARE",2)", then the model would execute the following line of SAS code:

absf_lung = Rand("CHISQUARE",2);

This is a powerful capability that can easily lead to errors if the input is not correct, so the user must be very careful when using this option.

The variable "n" indicates the number of points in the uncertainty vector or "cloud". If no uncertainty data are supplied, this should be zero. If "n" is zero in an uncertainty run, then the variability distribution is always used. Otherwise, each repetition of the uncertainty step randomly selects a number

US EPA ARCHIVE DOCUMENT

from 1 to n for each input record, and then uses the values at that location in the v1_un, v2_un, and v3_un vectors.

F.2.3 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 the probability vectors and the point distributions, may be correlated. The correlations file has three columns or SAS variables:

Table F-4. Column definitions for the correlations file.

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. 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 is empty if no correlations are requested.

Certain combinations of correlations are incompatible. Essentially, the correlation matrix must be positive definite (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.2.4 Contactmedia file

The "contactmedia" file contains a list of the various potential contact media in SHEDS. The current version of the model has 13 actual contact media, and 3 dummy media that are assigned zero concentrations. The SAS variables on this file are:

Table F-5. Column definitions for the contactmedia file.

Variable	Format	Purpose
medium	$6	name of contact medium as appearing in SAS code
medtype	$8	type of medium (must be "air", "matter", or "residue")
media	$24	category for this contact medium
medDescription	$40	longer description of contact medium

The default file contains the following 13 contact media:

Table F-6. Default contact media in the contactmedia file.

Medium	Medtype	Media	MedDescription
Tair	air	air (indoors)	Air in treated room
Tdust	matter	dust (indoors)	Dust in treated room
Thard	residue	smooth surfaces	Smooth surfaces in treated room
Tsoft	residue	textured surfaces	Textured surfaces in treated room
Uair	air	air (indoors)	Air in untreated room
Udust	matter	dust (indoors)	Dust in untreated room
Uhard	residue	smooth surfaces	Smooth surfaces in untreated room
Usoft	residue	textured surfaces	Textured surfaces in untreated room
Yair	air	air (outdoors)	Air outside residence
Ylawn	residue	lawn	Surface residues on lawn
Ysoil	matter	Soil (outdoors)	Soil in treated outdoor locations
Yveg	residue	vegetable Garden	Surface residues in vegetable garden
pet	residue	Pet	Residue on pet fur

The list of contact media in SHEDS is flexible. Media may be added or removed from the above list. However, at least two other input files must be modified when this is done. The "diarydetails" file gives contact probabilities and must refer to the correct list of media. The "appmedia" file gives the rules for determining concentration in various media after applications occur. Here, too, the list of affected media must match the list on the "contactmedia" file. With these provisos, SHEDS can be run with an altered list of media, without any changes to the code.

F.2.5 ModelDates file

The ModelDates file contains the rules for randomly assigning application dates to each individual. This file is only used if the ModelDates option is selected when the model is run. The file contains one record for each application type. The SAS variables on this file are:

Table F-7. Column definitions for the modeldates file.

Variable	Format	Purpose
appsymbol	$7	name of application type as appearing in SAS code
apploc	$8	location, set to either "Indoor" or "Outdoor"
apppriority	num	priority order for setting application dates
appdescrip	$40	description of application type
requirement	$40	logical condition necessary for application to occur
use	num	flag for including (=1) or excluding (=0) this record
user_pr	num	probability this application type is active, per individual
numapps_pr	$80	list of probabilities for number of applications per year
month_pr	$120	list of probabilities of application during each month
weekday_pr	$60	list of probabilities of application on each day of the week

F.2.6 UserDates file

Like ModelDates, the UserDates file also lists the possible application types, but is used when the UserDates option is selected. It contains the following SAS variables:

Table F-8. Column definitions for the userdates file.

Variable	Format	Purpose
appsymbol	$7	name of application type as appearing in SAS code
apploc	$8	location, set to either "Indoor" or "Outdoor"
appdescrip	$40	description of application type
use	num	flag for including (=1) or excluding (=0) this record
appdates	$200	list of day numbers on which this application occurs
apphour	num	hour of the day when application occurs

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 in appdates, which results 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. The apphour is a number between 1 and 24 inclusive, with 1 indicating midnight and 24 indicating 11 p.m.

F.2.7 AppMedia file

There are two versions of the AppMedia file, one (appmedia_decay) used with the decay-dispersion option and the other (appmedia_postapp) used with the post-application option. Each file contains one record for each combination of application type, affected contact medium, and type of input variable. However, unaffected contact media do not have to be listed. Two versions of this file are used because the required input variables are different.

Table F-9. SAS variables in both appmedia files.

Column name	Format	Purpose
description	$90	description of this application type
appsymbol	$7	name of application type as appearing in SAS code
medsymbol	$6	name of affected contact medium
varname	$8	input variable specified by this distribution
units	$16	the units for this input variable
use	num	flag for including (=1) or excluding (=0) this record
form	$16	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)
minval	num	the minimum allowed value for this variable
maxval	num	the maximum allowed value for this variable
notes	$200	user documentation

For the decay-dispersion method, the required input variables depend on the application type. Indoor applications require 12 variables each: the initial concentration in each of the 4 media in the treated room (air, dust, smooth surfaces, textured surfaces), the decay rate for each medium, and 4 untreated/treated area ratios (one for each medium). Outdoor applications require 4 inputs each: the initial soil and residue concentrations, and the decay rate for each. The default input file is based on outdoor air not being affected by applications. This could be changed by adding two more records to the file, for the initial outdoor air concentration and decay rate. The outdoor concentrations are assumed to disperse to non-contact media, so the dispersion does not have to be quantified (apart from its implied

US EPA ARCHIVE DOCUMENT

inclusion in the decay rate). The pet consists of just one affected medium, so only two inputs are required for each pet application type, the initial concentration and the decay rate.

For the post-application method, four inputs are required for every affected contact medium. Since there are 8 such media for indoor applications (4 in each of the treated and untreated areas), 32 inputs are needed for each indoor application type. Outdoor applications affect only 2 media, so 8 inputs suffice. The pet application type affect only one contact medium each, so 4 distributions are enough. In every case, the four distributions correspond to: day of application, 1-7 days after application, 8-30 days after application, and 31 or more days after application.

F.2.8 Diarydetails file

This file provides rules for determining when the simulated individual comes into contact with the various pesticide-containing contact media. The diarydetails file has a similar layout to the distributions file (copies of these files are later merged during the model run). Three records on this file give the distributions for the probability of being in a treated room, and untreated room, or far from an application site, for 'in-home' diary events. These are sampled once per person, then the results are normalized by dividing the results by their sum, so the probabilities total to unity. Every diary event that takes place 'in-home' makes a check to determine which location applies (treated, untreated, or neither).

The remaining records on this file give the distributions for contacting various media, given the location (treated room, untreated room, or outdoors in yard). These are also sampled once per person. There are multiple surfaces in each location, so a surface contact probability is obtained by sampling each of the distributions for surfaces and then dividing each value by the sum. This is not done for the soil and dust contact probabilities since there is only one soil/dust medium in each location, resulting in a yes/no test for contact. Contact probabilities for air are not required since there is only one possibility for the air medium in each location (Tair in the treated room, Uair in the untreated room, Yair outside in yard, and no contact elsewhere).

Table F-10. Contact probability definitions.

Input Variable	Description
probT_in_home	Probability of being in treated room while in home
probU_in_home	Probability of being in untreated room while in home
probN_in_home	Probability of being far from treated room while in home
probTR_RThard	Contact probability for smooth surface in treated room
probTR_RTsoft	Contact probability for textured surface in treated room
probTR_Rpet	Contact probability for pet while in treated room
probTR_Rnone	Contact probability for no surface while in treated room
probTM_MTdust	Contact probability for dust while in treated room
probUR_RUhard	Contact probability for smooth surface in untreated room
probUR_RUsoft	Contact probability for textured surface in untreated room
probUR_Rpet	Contact probability for pet while in untreated room
probUR_Rnone	Contact probability for no surface while in untreated room
probUM_MUdust	Contact probability for dust while in untreated room
probYR_RYlawn	Contact probability for lawn while outdoors at home
probYR_RYveg	Contact probability for garden while outdoors at home
probYR_Rpet	Contact probability for pet while outdoors at home
probYR_Rnone	Contact probability for no surface while outdoors at home
probYM_MYsoil	Contact probability for soil while outdoors at home

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 to age 20 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 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.

F.3.2 diarycohorts 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 22 age-gender cohorts and 8 season-daytype combinations, for a total of 176 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 diaryevents file

This is the large file (35 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-11. 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
chadact	$5	CHAD activity code
chadloc	$5	CHAD location code
micro	$8	SHEDS microenvironment
io	$1	“I”=indoors, “O”=outdoors
basenum	num	activity category for selecting METS distribution
add	$3	if “age” or “occ”, then METS depends on age or occup.
special	$1	“Z”=sleeping event, “B”=bath/shower event

The version of the diaryevents file in the SHEDS default directory contains 805941 records (diary events), covering 21114 diary days.

F.3.4 diarylevels file

This file contains the probability matrix for selecting diaries in various ranges of the spectrum, based on the chosen tendency for the individual. In SHEDS-Multimedia 3, the user may select both the variable and the number of levels to be used. An example of this input file is

Table F-12. Example diarylevels input file.

Obs	Diarylevel	High Tendency	Medium Tendency	Low Tendency
1	high	0.52	0.32	0.16
2	medium	0.30	0.38	0.32
3	low	0.18	0.30	0.52

Each simulated individual is assigned a tendency at random (with equal probability for each). Suppose a person is given the high tendency. Then there is a 52% chance that each diary drawn for this person belongs in the top third of the available pool for that cohort and day type, as ranked by outdoor time (which is the ranking variable currently used in the Diary PreProcessing module). Similarly, a high-tendency person has a 30% chance of being given a diary in the middle third of the pool, and an 18% chance of being given a diary in the lowest third.

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-13. 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 diarymets file

This file contains the activity-specific distributions for METS, which is the multiple of the basal metabolic rate needed to sustain the current activity. Essentially, this determines the inhalation rate (cubic meters of air per minute) and hence the rate at which chemical is inhaled. While most of the METS distributions depend only on the CHAD activity code, some are either age dependent or occupation dependent. In those cases, the age or occupation is taken from the personal profile of the simulated individual.

The “priority” variable on the diarymets file indicates the order into which this file will be sorted. This corresponds to the “basenum” (or the sum of “basenum” plus “add”) on the diaryevents file. For example, sleeping events have basenum=1, indicating that the METS distribution with priority=1 should be used (this corresponds to CHAD activity code “14500”).

F.3.7 diaryQArules file

This file contains a list of logical rules for accepting a CHAD diary for use in SHEDS Multimedia-3. 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)
- sleep>0

- LocUX<60
- ActUX<120

The first two of these rules are self-explanatory. The last three refer to the total time spent on the given diary day in various conditions. Rule 4 requires a non-zero amount of time spent sleeping. Rules 5 and 6 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 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. The above five tests are passed by 21114 of the 22968 CHAD diaries.

F.3.8 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.3.9 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 Tables for the One-Person, One-Day Simulation for Code Verification

The first table below (TableG-1) is 'Exposure_inputs', which lists in alphabetical order the variables input to the exposure module. This table has 30 rows, one for each of the 30 diary events. Each of the 50 columns represents one input variable and are spread over multiple pages. Some of these inputs retain the same value over the course of the day, while others may change hourly or change on every event.

US EPA ARCHIVE DOCUMENT

Table G-1. Inputs to exposure for one-day calculation.

	Absf_lung	Absr_dm	Absr_dr	Absr_gm	Absr_gr	Adherence	Air
1	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AUair
2	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AUair
3	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AUair
4	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AUair
5	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AUair
6	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AUair
7	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AUair
8	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AUair
9	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	ATair
10	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AUair
11	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	ATair
12	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	ATair
13	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	Anone
14	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AYair
15	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	Anone
16	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	ATair
17	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	ATair
18	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AYair
19	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AYair
20	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AUair
21	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	Anone
22	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	Anone
23	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AYair
24	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AYair
25	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AUair
26	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	ATair
27	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	Anone
28	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AUair
29	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AUair
30	1	0.2360227506	0.03	0.1004152252	0.493773219	0.1691767658	AUair

US EPA ARCHIVE DOCUMENT

Table G-1. Inputs to exposure for one-day calculation. (continued)

	Age	Bioavm	bioavr	bodysurf	bva	contactb	contacth	day
1	7	1	1	9394.8821629	0.1865811178	0.6644172234	1.3523020354	30
2	7	1	1	9394.8821629	0.1865811178	0.5485745472	1.1444942472	30
3	7	1	1	9394.8821629	0.1865811178	0.921664891	1.0774086166	30
4	7	1	1	9394.8821629	0.1865811178	0.3820707732	0.9548712872	30
5	7	1	1	9394.8821629	0.1865811178	0.6031846237	1.2190601488	30
6	7	1	1	9394.8821629	0.1865811178	0.5687310473	1.1251017717	30
7	7	1	1	9394.8821629	0.1865811178	0.6643221303	1.1090786708	30
8	7	1	1	9394.8821629	0.1865811178	0.6386863814	1.2640304912	30
9	7	1	1	9394.8821629	0.1865811178	0.6630644286	0.8777008682	30
10	7	1	1	9394.8821629	0.1865811178	0.6630644286	0.8777008682	30
11	7	1	1	9394.8821629	0.1865811178	0.3815076844	0.7391864303	30
12	7	1	1	9394.8821629	0.1865811178	0.2492349487	0.9577883555	30
13	7	1	1	9394.8821629	0.1865811178	0.397728873	1.1830762918	30
14	7	1	1	9394.8821629	0.1865811178	0.397728873	1.1830762918	30
15	7	1	1	9394.8821629	0.1865811178	0.817190445	1.267132498	30
16	7	1	1	9394.8821629	0.1865811178	0.817190445	1.267132498	30
17	7	1	1	9394.8821629	0.1865811178	0.2811584167	1.0401429858	30
18	7	1	1	9394.8821629	0.1865811178	0.2811584167	1.0401429858	30
19	7	1	1	9394.8821629	0.1865811178	0.218484714	1.1063733874	30
20	7	1	1	9394.8821629	0.1865811178	0.3393413418	1.1538400217	30
21	7	1	1	9394.8821629	0.1865811178	0.3393413418	1.1538400217	30
22	7	1	1	9394.8821629	0.1865811178	0.3393413418	1.1538400217	30
23	7	1	1	9394.8821629	0.1865811178	0.5557321904	1.3013771858	30
24	7	1	1	9394.8821629	0.1865811178	0.3715402273	1.2001373995	30
25	7	1	1	9394.8821629	0.1865811178	0.2167475332	1.3407833557	30
26	7	1	1	9394.8821629	0.1865811178	0.5804320074	1.0882988826	30
27	7	1	1	9394.8821629	0.1865811178	0.6002445524	1.1560686956	30
28	7	1	1	9394.8821629	0.1865811178	0.4376092504	1.3812610963	30
29	7	1	1	9394.8821629	0.1865811178	0.0951006094	1.1999141062	30
30	7	1	1	9394.8821629	0.1865811178	1.0376925441	1.2507148952	30

US EPA ARCHIVE DOCUMENT

Table G-1. Inputs to exposure for one-day calculation. (continued)

	dermaxb	dermaxh	duration	duration_hrs	f_uncloth	gender	handsurf
1	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
2	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
3	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
4	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
5	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
6	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
7	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
8	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
9	0.4713215214	0.8745446169	30	0.5	0.3025757674	F	391.45342345
10	0.4713215214	0.8745446169	30	0.5	0.3025757674	F	391.45342345
11	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
12	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
13	0.4713215214	0.8745446169	20	0.3333333333	0.3025757674	F	391.45342345
14	0.4713215214	0.8745446169	40	0.6666666667	0.3025757674	F	391.45342345
15	0.4713215214	0.8745446169	30	0.5	0.3025757674	F	391.45342345
16	0.4713215214	0.8745446169	30	0.5	0.3025757674	F	391.45342345
17	0.4713215214	0.8745446169	15	0.25	0.3025757674	F	391.45342345
18	0.4713215214	0.8745446169	45	0.75	0.3025757674	F	391.45342345
19	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
20	0.4713215214	0.8745446169	15	0.25	0.3025757674	F	391.45342345
21	0.4713215214	0.8745446169	15	0.25	0.3025757674	F	391.45342345
22	0.4713215214	0.8745446169	30	0.5	0.3025757674	F	391.45342345
23	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
24	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
25	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
26	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
27	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
28	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
29	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345
30	0.4713215214	0.8745446169	60	1	0.3025757674	F	391.45342345

US EPA ARCHIVE DOCUMENT

Table G-1. Inputs to exposure for one-day calculation. (continued)

	height	hm_fraction	hm_freq	hour	ingestion_indoor	ingestion_outdoor	io
1	126.04398493	0.053662002	9.1199751987	1	1.627226692	5.2487818764	I
2	126.04398493	0.053662002	5.6827881861	2	1.627226692	5.2487818764	I
3	126.04398493	0.053662002	2.6109326122	3	1.627226692	5.2487818764	I
4	126.04398493	0.053662002	8.8788187237	4	1.627226692	5.2487818764	I
5	126.04398493	0.053662002	12.186908072	5	1.627226692	5.2487818764	I
6	126.04398493	0.053662002	7.5074518321	6	1.627226692	5.2487818764	I
7	126.04398493	0.053662002	8.0726277436	7	1.627226692	5.2487818764	I
8	126.04398493	0.053662002	4.4816813883	8	1.627226692	5.2487818764	I
9	126.04398493	0.053662002	7.3515164366	9	1.627226692	5.2487818764	I
10	126.04398493	0.053662002	7.3515164366	9	1.627226692	5.2487818764	I
11	126.04398493	0.053662002	1.3068403716	10	1.627226692	5.2487818764	I
12	126.04398493	0.053662002	8.0782118682	11	1.627226692	5.2487818764	I
13	126.04398493	0.053662002	0.0785820961	12	1.627226692	5.2487818764	O
14	126.04398493	0.053662002	0.0785820961	12	1.627226692	5.2487818764	O
15	126.04398493	0.053662002	11.998464964	13	1.627226692	5.2487818764	I
16	126.04398493	0.053662002	11.998464964	13	1.627226692	5.2487818764	I
17	126.04398493	0.053662002	4.45769106	14	1.627226692	5.2487818764	I
18	126.04398493	0.053662002	0.0114539033	14	1.627226692	5.2487818764	O
19	126.04398493	0.053662002	0.0160497674	15	1.627226692	5.2487818764	O
20	126.04398493	0.053662002	1.3703199368	16	1.627226692	5.2487818764	I
21	126.04398493	0.053662002	1.3703199368	16	1.627226692	5.2487818764	I
22	126.04398493	0.053662002	1.3703199368	16	1.627226692	5.2487818764	I
23	126.04398493	0.053662002	0.5810095045	17	1.627226692	5.2487818764	O
24	126.04398493	0.053662002	0.2980541148	18	1.627226692	5.2487818764	O
25	126.04398493	0.053662002	4.5838523197	19	1.627226692	5.2487818764	I
26	126.04398493	0.053662002	4.3090907378	20	1.627226692	5.2487818764	I
27	126.04398493	0.053662002	4.9749062594	21	1.627226692	5.2487818764	I
28	126.04398493	0.053662002	16.838351227	22	1.627226692	5.2487818764	I
29	126.04398493	0.053662002	9.0226625903	23	1.627226692	5.2487818764	I
30	126.04398493	0.053662002	18.432886187	24	1.627226692	5.2487818764	I

US EPA ARCHIVE DOCUMENT

Table G-1. Inputs to exposure for one-day calculation. (continued)

	loc	matter	mets	metsnum	micro	month	object_ratio	om_area
1			0.8988469909	1	sleep	4	0.1	10.770061878
2			0.9587612449	1	sleep	4	0.1	10.770061878
3			0.932968246	1	sleep	4	0.1	10.770061878
4			0.8509135142	1	sleep	4	0.1	10.770061878
5			0.8030672191	1	sleep	4	0.1	10.770061878
6			0.8484525082	1	sleep	4	0.1	10.770061878
7			0.9164690478	1	sleep	4	0.1	10.770061878
8			0.8373588843	1	sleep	4	0.1	10.770061878
9	T	MTdust	1.7683676094	2	in_home	4	0.1	10.770061878
10	U	MUdust	2.5469042112	105	in_home	4	0.1	10.770061878
11	T	MTdust	1.4889374973	96	in_home	4	0.1	10.770061878
12	T	MTdust	1.5848968064	96	in_home	4	0.1	10.770061878
13			5.5025954845	71	out_oth	4	0.1	10.770061878
14			3.669822291	122	out_home	4	0.1	10.770061878
15			1.6866051326	2	in_home	4	0.1	10.770061878
16	T	MTdust	3.5202367774	18	in_home	4	0.1	10.770061878
17	T	MTdust	1.1224998806	27	in_home	4	0.1	10.770061878
18			3.3756344149	54	out_home	4	0.1	10.770061878
19			3.0297562272	54	out_home	4	0.1	10.770061878
20	U	MUdust	1.0844349168	27	in_home	4	0.1	10.770061878
21			2.0183554752	8	in_home	4	0.1	10.770061878
22			2.7845688783	105	in_home	4	0.1	10.770061878
23			3.4715920164	122	out_home	4	0.1	10.770061878
24			3.0748345865	122	out_home	4	0.1	10.770061878
25	U		1.9703609767	2	in_home	4	0.1	10.770061878
26	T		1.0181783128	4	in_home	4	0.1	10.770061878
27			1.0141645924	4	in_home	4	0.1	10.770061878
28			0.8143587651	1	sleep	4	0.1	10.770061878
29			0.9136262225	1	sleep	4	0.1	10.770061878
30			0.9335941203	1	sleep	4	0.1	10.770061878

US EPA ARCHIVE DOCUMENT

Table G-1. Inputs to exposure for one-day calculation. (continued)

	om_freq	om_transfer	remv_bath	remv_dry	remv_mouth	remv_wash
1	3.6983617687	0.3436606911	0.7569646266	0.02	0.1488417335	0.6080461864
2	3.1671875036	0.4988235416	0.7069967561	0.02	0.1991239961	0.6331881945
3	4.0628498304	0.3358044743	0.9264417148	0.02	0.1970584006	0.609437032
4	3.2550974871	0.3624097842	0.79262576	0.02	0.2529735522	0.5741257486
5	2.6435983524	0.3209726461	0.8038486621	0.02	0.1487225601	0.562781899
6	2.8083060131	0.1945103631	0.8794083146	0.02	0.1842450716	0.6107379794
7	1.8050757296	0.1470701067	0.8980142011	0.02	0.1780291472	0.6112209134
8	4.9322584215	0.410160597	0.8863498293	0.02	0.1967034311	0.4713087247
9	3.8842457937	0.3991666974	0.7416173635	0.02	0.2749755301	0.5474800419
10	3.8842457937	0.3991666974	0.7416173635	0.02	0.2749755301	0.5474800419
11	3.9480164717	0.1467936985	0.889863485	0.02	0.1367439061	0.6815952728
12	2.4224809415	0.3821706337	0.7550194187	0.02	0.1403372548	0.6916612455
13	1.3350437341	0.2183123125	0.8256186494	0.02	0.2279513834	0.6283204519
14	1.3350437341	0.2183123125	0.8256186494	0.02	0.2279513834	0.6283204519
15	1.8708297801	0.1221277418	0.7681593333	0.02	0.1522876537	0.554911897
16	1.8708297801	0.1221277418	0.7681593333	0.02	0.1522876537	0.554911897
17	1.8158803977	0.4810778052	0.7887896795	0.02	0.1991667477	0.6637137637
18	1.8158803977	0.4810778052	0.7887896795	0.02	0.1991667477	0.6637137637
19	4.9965145197	0.3026573104	0.7105363774	0.02	0.2062671828	0.6073462195
20	2.1622946603	0.4342272203	0.749770304	0.02	0.1056299595	0.6946743188
21	2.1622946603	0.4342272203	0.749770304	0.02	0.1056299595	0.6946743188
22	2.1622946603	0.4342272203	0.749770304	0.02	0.1056299595	0.6946743188
23	2.4698268461	0.1917364095	0.6941087425	0.02	0.197640246	0.6483360499
24	3.0080188038	0.4523044787	0.7609577029	0.02	0.1226737598	0.4824830572
25	4.4550073368	0.3780188074	0.8324829118	0.02	0.0837983583	0.6659258447
26	2.9251151839	0.2513111395	0.6615100876	0.02	0.1376153	0.6299409518
27	2.8382193902	0.2713338603	0.8642930303	0.02	0.1157596951	0.624894049
28	1.1815785878	0.3033313817	0.804110373	0.02	0.1834745904	0.5970706937
29	1.7429774668	0.2078922453	0.8729989719	0.02	0.064510614	0.5876976126
30	1.016349813	0.3424274125	0.6836096394	0.02	0.2029214599	0.602639821

US EPA ARCHIVE DOCUMENT

Table G-1. Inputs to exposure for one-day calculation. (continued)

	residue	special	start	transfer_dermal	washprob	weight	year
1		Z	0	0.0202995053	2.0515527507	25.894089166	2007
2		Z	60	0.0453483872	2.0515527507	25.894089166	2007
3		Z	120	0.0100040812	2.0515527507	25.894089166	2007
4		Z	180	0.0643016486	2.0515527507	25.894089166	2007
5		Z	240	0.0011392351	2.0515527507	25.894089166	2007
6		Z	300	0.1254194411	2.0515527507	25.894089166	2007
7		Z	360	0.071493742	2.0515527507	25.894089166	2007
8		Z	420	0.0410167348	2.0515527507	25.894089166	2007
9		W	480	0.0112494754	2.0515527507	25.894089166	2007
10	RUhard		510	0.0112494754	2.0515527507	25.894089166	2007
11	RThard		540	0.1679599325	2.0515527507	25.894089166	2007
12			600	0.0469615317	2.0515527507	25.894089166	2007
13			660	0.0001084798	2.0515527507	25.894089166	2007
14	RYveg	W	680	0.0001084798	2.0515527507	25.894089166	2007
15		W	720	0.0087250447	2.0515527507	25.894089166	2007
16	RTsoft		750	0.0087250447	2.0515527507	25.894089166	2007
17	RTsoft		780	0.0165099702	2.0515527507	25.894089166	2007
18	RYlawn		795	0.0165099702	2.0515527507	25.894089166	2007
19	RYlawn		840	0.0490193838	2.0515527507	25.894089166	2007
20	RUsoft		900	0.0080459485	2.0515527507	25.894089166	2007
21		B	915	0.0080459485	2.0515527507	25.894089166	2007
22			930	0.0080459485	2.0515527507	25.894089166	2007
23	RYlawn		960	0.0109855912	2.0515527507	25.894089166	2007
24	RYlawn	W	1020	0.0032803075	2.0515527507	25.894089166	2007
25	RUhard		1080	0.0074977774	2.0515527507	25.894089166	2007
26	RThard	W	1140	0.0075498563	2.0515527507	25.894089166	2007
27			1200	0.0006409991	2.0515527507	25.894089166	2007
28		Z	1260	0.0255606284	2.0515527507	25.894089166	2007
29		Z	1320	0.0755479011	2.0515527507	25.894089166	2007
30		Z	1380	0.0250837658	2.0515527507	25.894089166	2007

US EPA ARCHIVE DOCUMENT

The following equations are used for derived variables that may vary from one diary event to another. When the applicable equation depends on the status of other variables, the conditions are given in square brackets. Here, 'Z' means while sleeping, and '~Z' means while not sleeping. Also, 'W' means an event with hand washing, and 'B' means an event with a bath or shower. These conditions are indicated by the variable called 'special' in the exposure_inputs table. The 'Ldg' variables with a suffix of '0' (for example, LdgHr0) indicate that the value to be used is the final value of that variable from the prior diary event.

Table G-2. Verification equations for SHEDS Q/A

```
Hcr        = handsurf * contactH

Bcr        = bodysurf * contactB * f_uncloth

Ingestion  = ingestion_outdoor                          [if io='O']
           = ingestion_indoor                           [if io='I' and ~Z]
           = 0                                          [if Z]

ExpHr    = ConcR * duration_hrs * transfer_dermal * Hcr

ExpBr    = ConcR * duration_hrs * transfer_dermal * Bcr

ExpHm    = ConcM * duration_hrs * adherence/1000 * Hcr

ExpBm    = ConcM * duration_hrs * adherence/1000 * Bcr

ExpGr    = ConcR * object_ratio * bioavR * om_area * (1-(1-om_transfer)**(om_freq*duration_h rs))   [if ~Z]
         = 0                                                                                        [if Z]

ExpGm    = ConcM * duration_hrs * Ingestion/1000 * BioavM                                           [if ~Z]
         = 0                                                                                        [if Z]

ExpLa    = ConcA * duration_hrs * bva * Mets


Hmax     = dermaxH * handsurf

Hdrop    = Max (LdgHr0 + ExpHr + LdgHm0 + ExpHm – Hmax, 0)

HdropR   = Hdrop * (LdgHr0 + ExpHr) / (Hmax + Hdrop)

HdropM   = Hdrop * (LdgHm0 + ExpHm) / (Hmax + Hdrop)


Bmax     = dermaxB * bodysurf * f_uncloth;

Bdrop    = Max (LdgBr0 + ExpBr + LdgBm0 + ExpBm – Bmax, 0)

BdropR   = Bdrop * (LdgBr0 + ExpBr) / (Bmax + Bdrop)

BdropM   = Bdrop * (LdgBm0 + ExpBm) / (Bmax + Bdrop)
```


US EPA ARCHIVE DOCUMENT

F_dr = absr_dr * duration_hrs / 24

F_dm = absr_dm * duration_hrs / 24

F_gr = absr_gr * duration_hrs / 24

F_gm = absr_gm * duration_hrs / 24

AbsHr = (ldgHr0 + ExpHr – HdropR) * F_dr

AbsHm = (ldgHm0 + ExpHm – HdropM) * F_dm

AbsBr = (ldgBr0 + ExpBr – BdropR) * F_dr

AbsBm = (ldgBm0 + ExpBm – BdropM) * F_dm

AbsGr = (ldgGr0 + ExpGr) * F_gr

AbsGm = (ldgGm0 + ExpGm) * F_gm

AbsLa = ExpLa * absf_lung

AbsTot = AbsHr + AbsBr + AbsHm + AbsBm + AbsGr + AbsGm + AbsLa

Htm_remain = (1-remv_mouth)**(hm_freq * duration_hrs) [if ~Z]
= 1 [if Z]

HtmR = (LdgHr0 + ExpHr – HdropR – AbsHr)/2 * hm_fraction * (1-htm_remain)

RemainH = 1 [if Z]
= 1 – remv_wash [if W]
= 1 - remv_bath [if B]
= 1 – remv_dry * duration_hrs [otherwise]

RemainB = remainH [if Z or B]
= 1 – remv_dry * duration_hrs [otherwise]

LdgHr = (LdgHr0 + ExpHr – HdropR – AbsHr – HtmR) * RemainH

LdgHm = (LdgHm0 + ExpHm – HdropM – AbsHm) * RemainH

LdgBr = (LdgBr0 + ExpBr – BdropR – AbsBr) * RemainB

LdgBm = (LdgBm0 + ExpBm – BdropM – AbsBm) * RemainB

LdgGr = (LdgGr0 + ExpGr – AbsGr + HtmR * BioavR)

LdgGm = (LdgGm0 + ExpGm – AbsGm)

LdgLa = 0

All of the above variables (with four exceptions) are found on the file 'expo_ug' in the \output directory under the job name. The exceptions are HdropR, HdropM, BdropR, and BdropM, which are evaluated and used implicitly in the code without writing them to an output file. For aid in performing checks of these equations, these four variables were reconstructed and added to the table below.

US EPA ARCHIVE DOCUMENT

Each of the exposure equations may be "hand-checked" using the values in TableG-1 and Table G-3. Select any event (any row in the Tables) and use the values for the variables on that event, with the exception of the variables 'LdgHr0', 'LdgHm0', 'LdgBr0', 'LdgBm0', 'LdgGr0', and 'LdgGm0', which use the value from the prior event. All of these 'Ldg...0' variables are assumed to be zero on the first event of this diary. As an example, for the first event the equation 'Hcr = handsurf * contactH' evaluates numerically as 529.3632613 = 391.45342345 * 1.3523020354, which is correct within round-off error.

US EPA ARCHIVE DOCUMENT

Table G-3 contains the values of the derived exposure variables on each of the 30 diary events. These variables are listed in the same order as in the equations presented above.

Table G-3. Output variables from the exposure calculations

	ConcR	ConcM	ConcA	Hcr	Bcr	Ingestion	ExpHr
1	0	0	1	529.3632613	1888.7147095	0	0
2	0	0	1	448.01619119	1559.4129413	0	0
3	0	0	1	421.75529141	2619.9833109	0	0
4	0	0	1	373.78763431	1086.0987103	0	0
5	0	0	1	477.20526863	1714.6510224	0	0
6	0	0	1	440.42494028	1616.7110918	0	0
7	0	0	1	434.15264255	1888.4443919	0	0
8	0	0	1	494.80906314	1815.5705795	0	0
9	0	20	2	343.57900963	1884.8691689	1.627226692	0
10	5	3	1	343.57900963	1884.8691689	1.627226692	9.6627090685
11	10	20	2	289.35705871	1084.4980383	1.627226692	486.00392039
12	0	20	2	374.92953071	708.49113654	1.627226692	0
13	0	0	0	463.11926463	1130.6094218	5.2487818764	0
14	0	0	0	463.11926463	1130.6094218	5.2487818764	0
15	0	0	0	496.02335429	2322.9975979	1.627226692	0
16	5	20	2	496.02335429	2322.9975979	1.627226692	10.819564847
17	5	20	2	407.16753269	799.23881946	1.627226692	8.4029048056
18	30	0	0	407.16753269	799.23881946	5.2487818764	151.2522865
19	30	0	0	433.09365013	621.07856122	5.2487818764	636.89951571
20	2	3	1	451.67462661	964.63330755	1.627226692	1.8170753849
21	0	0	0	451.67462661	964.63330755	1.627226692	0
22	0	0	0	451.67462661	964.63330755	1.627226692	0
23	30	0	0	509.42855459	1579.7597137	5.2487818764	167.89121546
24	30	0	0	469.79789364	1056.1639099	5.2487818764	46.232446908
25	5	0	1	524.85423469	616.14034048	1.627226692	19.676200967
26	10	0	2	426.01832332	1649.9729863	1.627226692	32.163771193
27	0	0	0	452.54704862	1706.2933884	1.627226692	0
28	0	0	1	540.69938483	1243.9759222	0	0
29	0	0	1	469.71048472	270.33904817	0	0
30	0	0	1	489.59662749	2949.8109062	0	0

US EPA ARCHIVE DOCUMENT

Table G-3. Output variables from the exposure calculations (continued)

	ExpBr	ExpHm	ExpBm	ExpGr	ExpGm	ExpLa
1	0	0	0	0	0	0.1677078763
2	0	0	0	0	0	0.1788867448
3	0	0	0	0	0	0.1740742582
4	0	0	0	0	0	0.1587643946
5	0	0	0	0	0	0.1498371794
6	0	0	0	0	0	0.1583052174
7	0	0	0	0	0	0.1709958194
8	0	0	0	0	0	0.1562353566
9	0	0.5812558565	3.1887606994	0	0.0162722669	0.3299440052
10	53.009473513	0.0871883785	0.4783141049	3.3828594464	0.00244084	0.2376021173
11	1821.5221726	0.9790498271	3.6694374125	5.0154177812	0.0325445338	0.5556152452
12	0	1.2685873082	2.3972047815	0	0.0325445338	0.5914236355
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15	0	0	0	0	0	0
16	50.670644706	0.8391562684	3.9299722057	0.617725016	0.0162722669	0.6568097129
17	16.494261398	0.3444164316	0.6760631929	1.3869356449	0.0081361335	0.1047186412
18	296.89670516	0	0	19.087055606	0	0
19	913.34665087	0	0	26.975407277	0	0
20	3.8806949413	0.0573096394	0.1223951574	0.5708225144	0.00122042	0.0505837697
21	0	0	0	0	0	0
22	0	0	0	0	0	0
23	520.63783247	0	0	13.211151288	0	0
24	103.9362725	0	0	27.027431252	0	0
25	23.098415449	0	0	4.7357026098	0	0.3676321535
26	124.57058934	0	0	6.1511972747	0	0.3799456955
27	0	0	0	0	0	0
28	0	0	0	0	0	0.1519439687
29	0	0	0	0	0	0.1704654018
30	0	0	0	0	0	0.1741910345

US EPA ARCHIVE DOCUMENT

Table G-3. Output variables from the exposure calculations (continued)

	Hmax	Hdrop	HdropR	HdropM	Bmax	Bdrop
1	342.34348425	0	0	0	1339.8085705	0
2	342.34348425	0	0	0	1339.8085705	0
3	342.34348425	0	0	0	1339.8085705	0
4	342.34348425	0	0	0	1339.8085705	0
5	342.34348425	0	0	0	1339.8085705	0
6	342.34348425	0	0	0	1339.8085705	0
7	342.34348425	0	0	0	1339.8085705	0
8	342.34348425	0	0	0	1339.8085705	0
9	342.34348425	0	0	0	1339.8085705	0
10	342.34348425	0	0	0	1339.8085705	0
11	342.34348425	154.36548556	153.95439444	0.4110911175	1339.8085705	541.4267251
12	342.34348425	0	0	0	1339.8085705	0
13	342.34348425	0	0	0	1339.8085705	0
14	342.34348425	0	0	0	1339.8085705	0
15	342.34348425	0	0	0	1339.8085705	0
16	342.34348425	0	0	0	1339.8085705	0
17	342.34348425	0	0	0	1339.8085705	0
18	342.34348425	0	0	0	1339.8085705	282.3180473
19	342.34348425	511.98869564	511.11378092	0.8749147199	1339.8085705	891.95256238
20	342.34348425	0	0	0	1339.8085705	0
21	342.34348425	0	0	0	1339.8085705	0
22	342.34348425	0	0	0	1339.8085705	0
23	342.34348425	0	0	0	1339.8085705	0
24	342.34348425	0	0	0	1339.8085705	0
25	342.34348425	0	0	0	1339.8085705	0
26	342.34348425	0	0	0	1339.8085705	0
27	342.34348425	0	0	0	1339.8085705	0
28	342.34348425	0	0	0	1339.8085705	0
29	342.34348425	0	0	0	1339.8085705	0
30	342.34348425	0	0	0	1339.8085705	0

US EPA ARCHIVE DOCUMENT

Table G-3. Output variables from the exposure calculations (continued)

	BdropR	BdropM	F_dr	F_dm	F_gr	F_gm
1	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
2	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
3	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
4	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
5	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
6	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
7	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
8	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
9	0	0	0.000625	0.0049171406	0.0102869421	0.0020919839
10	0	0	0.000625	0.0049171406	0.0102869421	0.0020919839
11	539.33538596	2.0913391439	0.00125	0.0098342813	0.0205738841	0.0041839677
12	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
13	0	0	0.0004166667	0.0032780938	0.0068579614	0.0013946559
14	0	0	0.0008333333	0.0065561875	0.0137159228	0.0027893118
15	0	0	0.000625	0.0049171406	0.0102869421	0.0020919839
16	0	0	0.000625	0.0049171406	0.0102869421	0.0020919839
17	0	0	0.0003125	0.0024585703	0.005143471	0.0010459919
18	280.33336822	1.9846790863	0.0009375	0.007375711	0.0154304131	0.0031379758
19	888.27203901	3.6805233677	0.00125	0.0098342813	0.0205738841	0.0041839677
20	0	0	0.0003125	0.0024585703	0.005143471	0.0010459919
21	0	0	0.0003125	0.0024585703	0.005143471	0.0010459919
22	0	0	0.000625	0.0049171406	0.0102869421	0.0020919839
23	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
24	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
25	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
26	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
27	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
28	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
29	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677
30	0	0	0.00125	0.0098342813	0.0205738841	0.0041839677

US EPA ARCHIVE DOCUMENT

Table G-3. Output variables from the exposure calculations (continued)

	AbsHr	AbsHm	AbsBr	AbsBm	AbsGr	AbsGm
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0
9	0	0.0028581168	0	0.0156795848	0	0.0000340413
10	0.0060391932	0.0017157128	0.0331309209	0.0179544238	0.0347992791	0.0000390763
11	0.4267897356	0.008965873	1.6682917073	0.0508944182	0.175765058	0.0002141544
12	0.415771631	0.0211757905	1.6328822158	0.0729608161	0.2050585772	0.0003494237
13	0.1330822191	0.006849397	0.5327414303	0.0235994777	0.1100549975	0.0001159872
14	0.2642319564	0.0135628621	1.0579386499	0.0467306045	0.2193940007	0.0002316509
15	0.0735692473	0.0037559914	0.7927927758	0.0348181724	0.1634657516	0.0001732536
16	0.0389344098	0.0057897762	0.8239664332	0.0539711926	0.1885568683	0.0002069325
17	0.0215614178	0.003698619	0.4127629265	0.0282465242	0.1058451276	0.0001117601
18	0.2057612197	0.0110132343	1.2472404336	0.0694701589	0.6166846817	0.0003349296
19	0.4271980859	0.0057532076	1.6678500362	0.0543692321	1.3647782703	0.0004451715
20	0.1050901581	0.0015365739	0.409325197	0.0134903818	0.3372852046	0.0001121038
21	0.1044267927	0.0015251321	0.4071512964	0.0133899287	0.3372912177	0.0001119865
22	0.0521924396	0.0007613901	0.2036990144	0.0066846404	0.6745724222	0.0002237388
23	0.3129364147	0.0015001396	1.0538692616	0.0131705067	1.6104633168	0.0004465414
24	0.3630979647	0.0014556791	1.1614212272	0.0127801646	2.149964621	0.0004446731
25	0.2120770692	0.0007459301	1.1888424699	0.0126544808	2.2092901425	0.0004428126
26	0.2459399289	0.0007238225	1.3193225252	0.0122794325	2.3213020056	0.0004409599
27	0.0897482486	0.0002652229	1.317673372	0.0121586731	2.3247048733	0.0004391149
28	0.0867644383	0.0002573623	1.2897057547	0.0117983192	2.2949968609	0.0004372777
29	0.0866559827	0.0002548313	1.2880936225	0.0116822912	2.2477798615	0.0004354481
30	0.0865476627	0.0002523253	1.2864835055	0.0115674043	2.201534299	0.0004336262

US EPA ARCHIVE DOCUMENT

Table G-3. Output variables from the exposure calculations (continued)

	AbsLa	AbsTot	Htm_remain	HtmR	RemainH	RemainB
1	0.1677078763	0.1677078763	1	0	1	1
2	0.1788867448	0.1788867448	1	0	1	1
3	0.1740742582	0.1740742582	1	0	1	1
4	0.1587643946	0.1587643946	1	0	1	1
5	0.1498371794	0.1498371794	1	0	1	1
6	0.1583052174	0.1583052174	1	0	1	1
7	0.1709958194	0.1709958194	1	0	1	1
8	0.1562353566	0.1562353566	1	0	1	1
9	0.3299440052	0.3485157482	0.3066833219	0	0.4525199581	1
10	0.2376021173	0.3312807235	0.3066833219	0.1796370471	0.99	0.99
11	0.5556152452	2.8865361916	0.8251724507	1.5995856162	0.98	0.98
12	0.5914236355	2.9396220898	0.2947735619	6.2858945738	0.98	0.98
13	0	0.8064435087	0.9932463101	0.057853318	0.9933333333	0.9933333333
14	0	1.6020897246	0.9865382326	0.1144309443	0.3716795481	1
15	0	1.0685751921	0.3711471421	1.9848637071	0.445088103	1
16	0.6568097129	1.7682353255	0.3711471421	1.0504320715	0.99	0.99
17	0.1047186412	0.6769450165	0.7807376404	0.4057817533	0.995	0.995
18	0	2.1505046578	0.9980938633	0.0112143943	0.985	0.985
19	0	3.5203940035	0.9962992341	0.0338925755	0.98	0.98
20	0.0505837697	0.9174233888	0.9624779257	0.3384542714	0.995	0.995
21	0	0.8638963541	0.9624779257	0.3363178311	0.250229696	0.250229696
22	0	0.9381336455	0.9263637575	0.1648862934	0.99	0.99
23	0	2.9923861809	0.879908594	0.8056597861	0.98	0.98
24	0	3.6891643296	0.9617427866	0.2977973319	0.5175169428	1
25	0.3676321535	3.9916850587	0.6695340338	1.5024641107	0.98	0.98
26	0.3799456955	4.27995437	0.5283600338	2.4866994431	0.3700590482	1
27	0	3.7449895048	0.5422409154	0.8807377444	0.98	0.98
28	0.1519439687	3.8359039818	1	0	1	1
29	0.1704654018	3.8053674392	1	0	1	1
30	0.1741910345	3.7610098576	1	0	1	1

US EPA ARCHIVE DOCUMENT

Table G-3. Output variables from the exposure calculations (continued)

	LdgHr	LdgHm	LdgBr	LdgBm	LdgGr	LdgGm	LdgLa
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	0	0.2617365209	0	3.1730811146	0	0.0162382256	0
10	9.3822625	0.3437370947	52.446579167	3.5971063877	3.5276972143	0.0186399893	0
11	332.61730484	0.8846753327	1306.3057726	5.0218240333	9.9669355538	0.0509703688	0
12	319.39732586	2.0894451134	1278.5794326	7.1991466387	16.04777155	0.083165479	0
13	317.07834772	2.068711745	1269.5263799	7.12771018	15.995569871	0.0830494917	0
14	117.71079574	0.7638568081	1268.4684413	7.0809795755	15.890606815	0.0828178408	0
15	51.475490758	0.3383118306	1267.6756485	7.0461614031	17.71200477	0.0826445872	0
16	60.593632233	1.1599615395	1304.3471035	10.812940792	19.191604989	0.0987099217	0
17	68.226347898	1.4931759553	1313.8264589	11.403453673	20.87847726	0.106734295	0
18	215.9727339	1.4599302802	1309.2054171	9.2090648618	39.360062579	0.1063993654	0
19	334.47143047	0.5676771057	1305.9599354	5.3646888168	65.004584161	0.1059541939	0
20	334.16573662	0.6203329204	1302.8841486	5.4462256244	65.576575743	0.1070625101	0
21	83.507903293	0.1548440847	325.91842307	1.3594568245	65.575602356	0.1069505236	0
22	82.457916314	0.1525418676	322.45757682	1.3392444623	65.065916227	0.1067267848	0
23	244.24592486	0.1480208935	825.20070923	1.2995524764	77.472263985	0.1062802434	0
24	149.98545439	0.0758499816	927.9755605	1.2867723119	102.64752795	0.1058355703	0
25	164.58817189	0.0736019706	930.88743081	1.2486354744	106.67640453	0.1053927577	0
26	71.798598862	0.0269692181	1054.1386976	1.2363560419	112.99299924	0.1049517978	0
27	69.411550611	0.0261699153	1031.7646038	1.1997134215	111.54903211	0.1045126829	0
28	69.324786173	0.025912553	1030.474898	1.1879151022	109.25403525	0.1040754052	0
29	69.23813019	0.0256577217	1029.1868044	1.176232811	107.00625539	0.103639957	0
30	69.151582528	0.0254053964	1027.9003209	1.1646654067	104.80472109	0.1032063308	0

US EPA ARCHIVE DOCUMENT

United States
Environmental Protection
Agency

(8101R)
Washington, DC 20460

Official Business
Penalty for Private Use
$300

EPA 600/R-08/118
September 2008
www.epa.gov

US EPA ARCHIVE DOCUMENT

Recycled/Recyclable Printed with vegetable-based ink on paper that contains a minimum of 50% post-consumer fiber content processed chlorine free