DATA COLLECTION FOR THE HAZARDOUS WASTE IDENTIFICATION RULE

SECTION 8.0 HUMAN EXPOSURE FACTORS

Work Assignment Managers and Technical Direction:	Stephen M. Kroner David A. Cozzie U.S. Environmental Protection Agency Office of Solid Waste Washington, DC 20460
Prepared by:	Center for Environmental Analysis Center for Environmental Measurements and Quality Assurance Statistics Research Division Research Triangle Institute 3040 Cornwallis Road Research Triangle Park, NC 27709-2194 Under Contract No. 68-W-98-085

U.S. Environmental Protection Agency
Office of Solid Waste
Washington, DC 20460

October 1999

ACKNOWLEDGMENTS

This work was performed by the Research Triangle Institute (RTI) under U.S. Environmental Protection Agency (EPA) contract 68-W-98-085 with the Office of Solid Waste. Stephen Kroner, the U.S. EPA Work Assignment Manager, provided overall technical direction and review throughout this work. Terry Pierson, the RTI Work Assignment Leader, along with Robert Truesdale, leader of the data collection task, provided day-to-day management and technical direction at RTI. Larry Myers developed, implemented, and documented the statistical methodologies used to fit and parameterize distributions for the human exposure factors. Susan Wolf extracted the data used to fit the distributions, provided guidance and technical direction to the distribution-fitting work, and assisted with documentation. Jenny Lloyd entered and compiled the data for use in the model. Cindi Salmons was the quality assurance officer.

DISCLAIMER

The work presented in this document has been funded by the United States Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the Agency.

Table of Contents

Section **Page**

Appendices

Table of Contents (continued)

Appendices **Page**

List of Tables

8.0 Human Exposure Factors

Exposure factors are used in the human exposure module and the human risk module of the Hazardous Waste Identification Rule (HWIR) risk modeling system to calculate the dose of a chemical (in mg/kg of body weight/d) based on contact with contaminated media or food, the duration of that contact, and the body weight of the exposed individuals. This section describes the derivation and use of values for the human exposure factors used in the HWIR analysis.

The human exposure module calculates exposures to human receptors from media and food concentrations calculated by other modules. The human exposure module calculates exposures for two basic receptor types: residential receptors (residents and home gardeners) and farmers. Residential receptors also may be recreational fishers, in addition to being a resident or home gardener. Farmers may be beef farmers or dairy farmers, and either type of farmer also may be a recreational fisher. The subcategories within residential receptors and farmers differ in the particular exposures they incur. For example, a resident (only) differs from a home gardener in that home gardeners are exposed to contaminated produce, but residents are not. Within each of the two basic receptor types, the human exposure module calculates exposures for five age cohorts: infants (ages 0 to 1 year), children ages 1 to 5 years, children ages 6 to 11 years, children ages 12 to 19 years, and adults (ages 20 years or older).

The media inputs needed for the human exposure module include ambient air concentration (both vapor and particulate), soil concentration, groundwater concentration, exposed vegetable concentration, protected vegetable concentration, exposed fruit concentration, protected fruit concentration, root vegetable concentration, beef concentration, milk concentration, and fish filet concentration for trophic level 3 and trophic level 4 fish. For vegetables and fruits, the terms “exposed” and “protected” refer to whether the edible portion of the plant is exposed to the atmosphere.

Exposure to humans other than infants may occur through eight pathways:

Inhalation of ambient air
Inhalation of shower air
Ingestion of groundwater
Ingestion of soil
Ingestion of fruits and vegetables
Ingestion of beef
Ingestion of milk
Ingestion of fish.

However, not all receptors are exposed through all of these pathways. Residents are exposed through inhalation of ambient air, inhalation of shower air, ingestion of groundwater, and ingestion of soil. Home gardeners have the same exposures as residents, plus exposure through ingestion of fruits and vegetables. All farmers are exposed through inhalation of ambient air, inhalation of shower air, ingestion of groundwater, ingestion of soil, and ingestion of fruits and vegetables. In addition, beef farmers are exposed through ingestion of beef, and dairy farmers are exposed through ingestion of milk. Recreational fishers have the same exposures as one of the other receptor types, plus fish ingestion. Not all age cohorts are exposed through all pathways–shower exposures are calculated only for adults and children aged 12 to 19 years.

The generalized equation for calculating dose based on exposure through a single pathway (e.g., ingestion of contaminated drinking water or ingestion of contaminated produce) is the following:

$$Dose_i = \frac{C_i \bullet CR_i \bullet F_i}{BW} \tag{8-1}$$

where

$Dose_i$	=	mg/kg of body weight/d of contaminant taken into the body through exposure pathway i
C_i	=	concentration (in mg/kg, mg/L, or mg/m^3) of a contaminant in the media or food product associated with exposure pathway i
CR_i	=	contact rate (in L/d, g/d, or m^3/d) with the media or food product associated with exposure pathway i (e.g., ingestion rate of water per day or inhalation rate of air per day)
F_i	=	fraction of the total contact rate that is associated with the contaminated media or food (e.g., fraction of air breathed [m^3/d] or water ingested [L/d] that is contaminated)
BW	=	body weight for the exposed cohort (e.g., adult, infant, and so on).

Carcinogens include the additional dimension of the duration of the exposure averaged over a lifetime.

The human exposure factors used in the HWIR FRAMES system are based on national data for these factors provided in the *Exposure Factors Handbook* (EFH) (U.S. EPA, 1997a, 1997b, and 1997c). Where appropriate, distributions were developed for use in the HWIR analysis. This section explains the parameters for which data were obtained (Section 8.1), the data sources (Section 8.2), and the methods (Section 8.3) used to develop distributions or fixed values for all the exposure factors used in the analysis (Section 8.4). Uncertainties and issues associated with these variables are described in Section 8.5.

8.1 Parameters Collected

The exposure factors used in the HWIR risk analysis are shown in Table 8-1, along with the data source and whether they are represented by a distribution or as a fixed value. These inputs address inhalation and ingestion exposure from contact with media and various food items as well as duration of exposure for the receptor populations or types addressed in the HWIR analysis: residents, home gardeners, beef and dairy farmers, and recreational fishers.

Age is a relevant covariate (i.e., an important determinant) for most environmental exposure factors, and stratification on age in risk assessment simulations is commonly used to reduce variability. Although there are no universally recognized standardized definitions of age groups, this analysis uses the following: child1 (0 to 1 year old), child2 (1 to 5 years old), child3 (6 to 11 years old), child4 (12 to 19 years old), and adult (older than 19 years of age). These age groups were selected because a majority of the data in the EFH were provided in a similar manner, and the use of these age groups in this analysis (for all receptor types) reduced the need to manipulate the data sets. In addition, other human risk analyses performed by U.S. Environmental Protection Agency's (EPA's) Office of Solid Waste (OSW) have used these same age cohort definitions.

Site-specific and regional data sets are not available for many of the human exposure inputs. Although food consumption rates and exposure duration data are grouped by four regions (Northeast, Midwest, South, and West) in the EFH, all age groups were combined when these data were presented by region. The unavailability of similar regional distributions for other human exposure inputs in addition to the loss of age-specific data indicated that deriving regional distributions was not feasible. Therefore, all human exposure model inputs were collected and processed on a national basis.

The HWIR human exposure parameters either are characterized by distributions (stochastic variables) or are fixed values (constants). Table 8-1 shows this breakdown; national distributions were developed for all factors with data that could be used to derive distributions. A few parameters were fixed based on central tendency values from the best available source, either because limited variability was expected or because available data were not adequate to generate national distributions.

8.2 Data Sources

The following documents are the primary data sources for the exposure factors used for HWIR:

- U.S. Environmental Protection Agency (EPA). 1997a. *Exposure Factors Handbook, Volume I, General Factors.* Office of Research and Development, Washington, DC. August.

- U.S. Environmental Protection Agency (EPA). 1997b. *Exposure Factors Handbook, Volume II, Food Ingestion Factors.* Office of Research and Development, Washington, DC. August.

Table 8-1. Input Parameters and Data Sources: HWIR Human Exposure Factors

Abbreviation[a]	Parameter	Data Source
Distributed (stochastic) variables		
bodywt	Body weight (adult, child1-4)	U.S. EPA (1997a)
inhal	Inhalation rate (adult, child1-4)	U.S. EPA (1997a)
soilIng	Ingestion rate: soil (adult, child2-4)	U.S. EPA (1997a)
drinkH2O	Ingestion rate: drinking water (adult, child1-4)	U.S. EPA (1997a)
bmilk	Breast milk consumption (child1)	U.S. EPA (1997b)
expveg	Consumption rate for gardener: exposed vegetables (adult, child2-4)	U.S. EPA (1997b)
	Consumption rate for farmer: exposed vegetables (adult, child2-4)	U.S. EPA (1997b)
rootveg	Consumption rate for gardener: root vegetables (adult, child2-4)	U.S. EPA (1997b)
	Consumption rate for farmer: root vegetables (adult, child2-4)	U.S. EPA (1997b)
proveg	Consumption rate for gardener: protected vegetables (adult, child2-4)	U.S. EPA (1997b)
	Consumption rate for farmer: protected vegetables (adult, child2-4)	U.S. EPA (1997b)
expfruit	Consumption rate for gardener: exposed fruit (adult, child2-4)	U.S. EPA (1997b)
	Consumption rate for farmer: exposed fruit (adult, child2-4)	U.S. EPA (1997b)
profruit	Consumption rate for gardener: protected fruit (adult, child2-4)	U.S. EPA (1997b)
	Consumption rate for farmer: protected fruit (adult, child2-4)	U.S. EPA (1997b)
fish	Consumption rate for recreational fisher: fish (adult, child2-4)	U.S. EPA (1997b)
beef	Consumption rate for farmer: beef (adult, child2-4)	U.S. EPA (1997b)
milk	Consumption rate for farmer: milk (adult, child2-4)	U.S. EPA (1997b)
showerT	Shower contact time	U.S. EPA (1997c)
cumTroom	Total time in shower and bathroom	U.S. EPA (1997c)
Fixed variables (constants)		
-	Exposure frequency (adult, child1-4)	EPA policy
-	Exposure duration (adult, child1-4)	U.S. EPA (1997c)
-	Fraction contaminated: soil	EPA policy
-	Fraction contaminated: drinking water	EPA policy
-	Fraction contaminated for recreational fisher (fish)	U.S. EPA (1997b)
-	Fraction homegrown for gardener (exposed vegetables, root vegetables, protected vegetables, exposed fruit, protected fruit)	U.S. EPA (1997b)
-	Fraction homegrown for farmer (exposed vegetables, root vegetables, protected vegetables, exposed fruit, protected fruit)	U.S. EPA (1997b)
-	Fraction contaminated (home-raised) for farmer (beef, dairy)	U.S. EPA (1997b)
-	Event frequency–showering	U.S. EPA (1997c)
-	Fraction of fat in maternal breast milk	U.S. EPA (1997b)
-	Fraction of T3 fish consumed	U.S. EPA (1997b)
-	Fraction of T4 fish consumed	U.S. EPA (1997b)
-	Human lifetime (used in carcinogenic risk calculation)	U.S. EPA (1997a)

[a]Abbreviations are used only for stochastic parameters.

- # U.S. Environmental Protection Agency (EPA). 1997c. *Exposure Factors Handbook*, *Volume III, Activity Factors*. Office of Research and Development, Washington, DC. August.

The EFH is the most important data source for most human exposure model inputs. It summarizes data on human behaviors and characteristics related to human exposure from relevant key studies and provides recommendations and associated confidence estimates on the values of exposure factors. By providing a consistent set of exposure factors, the EFH serves as a support document to EPA's *Guidelines for Exposure Assessment* (U.S. EPA, 1992), developed to promote consistency among various exposure assessment activities across different program offices. For many factors, percentile data that can be used to develop distributions are provided in the handbook.

EPA carefully reviews and evaluates data quality before inclusion in the EFH. Evaluation criteria include peer review, reproducibility, pertinence to the United States, currency, adequacy of the data collection period, validity of the approach, representativeness of the population, characterization of the variability, lack of bias in study design, and measurement error (U.S. EPA, 1997a, 1997b, 1997c).

8.3 Methodology

The general methodology for collecting human exposure data relied upon data provided by the EFH, which were used in one of three ways:

1. When EFH percentile data were adequate (most input variables), maximum likelihood estimation was used to fit selected parametric models (gamma, lognormal, Weibull, and generalized gamma) to the EFH data. The chi-square measure of goodness of fit was then used to choose the best distribution to assume for HWIR. Parameter uncertainty information (e.g., for averages, standard deviations) also was derived using the asymptotic normality of the maximum likelihood estimate or a regression approach.

2. For a few variable conditions when percentile data were not adequate for model fitting, models were selected on the basis of results for other age cohorts or, if no comparable information was available, by assuming lognormal as a default distribution and reasonable coefficients of variation (CVs).

3. Other variables for which data were not adequate for either 1 or 2 above were fixed at EFH-recommended mean values or according to established Agency policy.

The approach for developing national distributions (1 and 2) is described in Section 8.3.1, followed by discussions of the development of fixed parameters (Section 8.3.2) and quality assurance (QA) and quality control (QC) procedures (Section 8.3.3). Results by parameter are provided in Section 8.4.

8.3.1 National Distributions

For most variables for which national distributions were developed, exposure factor data from the EFH were analyzed to fit selected parametric models. Steps in this process include preparing data, fitting models, assessing fit, and preparing parameters to characterize distributional uncertainty in the HWIR model inputs.

Because the EFH data are always positive and almost always skewed to the right (i.e., have a long right tail),[1] three two-parameter probability models commonly used to characterize such data (gamma, lognormal, and Weibull) were selected. In addition, a three-parameter model (generalized gamma) was used that unifies them[2] and allows for a likelihood ratio test of the fit of the two-parameter models. However, only the two-parameter models were selected for use in the analysis because the three-parameter generalized gamma model did not significantly improve the goodness of fit over the two-parameter models in 58 of 59 cases at the 5 percent level of significance.

This simple setup constitutes a considerable improvement over the common practice of using a lognormal model in which adequate EFH data were available to support maximum likelihood estimation. However, in a few cases (soil ingestion, breast milk consumption, and inhalation rate), data were not adequate to fit a distribution, and the lognormal model was assumed as the default.

8.3.1.1 <u>Preparation of Percentile Data</u>. For many exposure factors, EFH data include sample sizes and estimates of the following parameters for specific receptor types and age groups: mean, standard deviation, standard error, and percentiles corresponding to a subset of the following probabilities–0.01, 0.02, 0.05, 0.10, 0.15, 0.25, 0.50, 0.75, 0.85, 0.90, 0.95, 0.98, and 0.99. These percentile data were used as a basis of fitting distributions where available. Although in no case are all of these percentiles actually provided for a single factor, seven or more are typically present in the EFH data. Therefore, using the percentiles is a fuller use of the available information than simply fitting based on the method of moments (e.g., selecting models that agree with the data mean and standard deviation).

For some factors, certain percentiles were not used in the fitting process because sample sizes were too small to justify their use. Percentiles were used only if at least one data point was in the tail of the distribution. For example, the 1st and 99th percentiles were used only when sample sizes exceeded 100; when sample size data were unavailable, the 5th through 95th percentiles were used.

If the EFH data repeated a value across several adjacent percentiles, only one value (the most central) was used in most cases. For example, for the time in the shower for those aged 12 to 19, the 1st, 2nd, and 5th percentiles were all equal to 5 min. This means that 1 percent of the data is below 5, 2 percent of the data are below 5, and 5 percent of the data are below 5, in

[1] In preliminary goodness-of-fit tests, the normal distribution fit best in only 3 cases out of 69.

[2] Gamma, Weibull, and lognormal are all special cases of the generalized gamma.

essence indicating that 5 percent of the data are below 5. In this case, only the 5th percentile (the most central or closest to the 50th) was used in the analysis. Similarly, because both the 98th and 99th percentiles are listed as 60 min, only the 98th value was used. In this case, the data summary indicates that 2 percent of the data are above 60 min and 1 percent of the data is above 60 min. The first statement implies the second, so the 99th percentile was dropped from the analysis.

The EFH does not use standardized age cohorts across exposure factors. Therefore, to obtain the percentiles for fitting the five standardized HWIR age cohorts (i.e., less than 1, 1 to 5, 6 to 11, 12 to 19, and more than 20), each EFH cohort-specific value for a given exposure factor was assigned to one of these five cohorts. When multiple EFH cohorts fit into a single HWIR cohort, the EFH percentiles were averaged within each cohort. If sample sizes were available, weighted averages were used, with weights proportional to sample sizes. If sample sizes were not available, equal weights were assumed (i.e., the percentiles were simply averaged).

Appendix 8A contains the final EFH data used in the analysis, developed by applying the methods described previously to the EFH data. All raw data used to develop each exposure factor distribution also are summarized by parameter in Section 8.4.

8.3.1.2 Statistical Methods and Data Analysis. The following statistical methods were used to fit distributions to the EFH percentile data, to determine the goodness of fit of these distributions, and to develop the statistics needed to describe the distributions in the HWIR modeling system:

- **Model Fitting**–Lognormal, gamma, Weibull, and generalized gamma distributions were fit to each factor data set using maximum likelihood estimation (Burmaster and Thompson, 1998).

- **Assessment of Fit**–When sample sizes were available, the goodness of fit was calculated for each of the four models using the chi-square test (Bickel and Doksum, 1977). For the two-parameter models, the goodness of fit was calculated against the generalized gamma model using the maximized log likelihood (likelihood ratio) test (Bickel and Doksum, 1977). When percentile data were available but sample sizes were unknown, a regression F-test for the goodness of fit against the generalized gamma model was used.

- **Distributions for Parameter Uncertainty**–For each of the two-parameter models, parameter uncertainty information was provided as parameter estimates for a bivariate normal distribution that could be used for simulating parameter values (Burmaster and Thompson, 1998). The information necessary for such simulations includes estimates of the two model parameters, their standard errors, and their correlation. To obtain this parameter uncertainty information, one can use the asymptotic normality of the maximum likelihood estimate (Burmaster and Thompson, 1998) when sample sizes are available, and a regression approach when sample sizes are not available (Jennrich and Moore, 1975; Jennrich and Ralston, 1979). In either case, uncertainty can be expressed as a bivariate normal distribution for the model parameters.

The chi-square statistic and the maximized log likelihood both were used to compare the 3 two-parameter models (gamma, lognormal, and Weibull) for each parameter. The two statistics agreed on which two-parameter model fit best in 61 of 69 cases. For testing goodness of fit, the chi-square test of absolute fit against the data is more stringent than the likelihood ratio test of relative fit against the generalized gamma model and was used for this analysis.

Each test required a known sample size, which was available in 59 of the 69 cases. If a *p*-value above 0.05 represents adequate fit, then adequate fit was obtained by 1 of the two-parameter models in 46 of 59 cases using the likelihood ratio test versus 30 of 59 cases for the chi-square test. Using the chi-square test at the 5 percent significance level, in only 1 of 59 cases did the generalized gamma model offer improved fit over the best two-parameter model. However, if the likelihood ratio test was used, then the generalized gamma model improved significantly on the best-fitting two-parameter model in 13 of 59 cases.

8.3.1.3 Analysis Results–Goodness of Fit and Parameterization of Distributions. Table 8-2 and Appendices 8B and 8C summarize the results of this statistical analysis. The tabulated results are applicable using any software tool capable of generating random variables from lognormal, gamma, and Weibull distributions, such as Crystal Ball or @RISK.

Table 8-2 contains the results from the analysis based on the chi-square goodness-of-fit test. Results from the likelihood ratio test are not shown but are available on request. In Table 8-2, M is the number of percentiles used in the fitting process. The ranking of the 3 two-parameter models based on the chi-square statistic is indicated by columns labeled First, Second, and Third. Values of the chi-square statistic also are shown for all three models. PGOF denotes the *p*-value for the chi-square test of goodness of fit. Small values of PGOF, such as PGOF <0.01, correspond to large chi-square values and tend to cast doubt on the fit of the model.

Appendix 8B identifies the models selected for use in HWIR and provides the estimated means and standard deviations for each of the two-parameter models fit to the exposure factors data. These quantities characterize the models for use in risk assessment. For instance, because the gamma model was the best-fitting model for the showerT variable in Table 8-2, we used the gamma distribution with a mean of 16.7 min and a standard deviation of 9.91 min for the HWIR model inputs.

Appendix 8C has been provided for the convenience of risk assessors who want to use the distributions with Crystal Ball Monte Carlo software or software requiring similar statistics. It contains parameter estimates for each model based on the parameterizations used in Appendix B of the Crystal Ball manual. For instance, columns labeled LOG MEAN and LOG SDEV contain the values of the lognormal parameters μ and σ (page 255 of the Crystal Ball manual). Similarly, the last four columns of Appendix 8C contain the values of the parameters α (shape) and β (scale) for each of the gamma and Weibull distributions. Note that in HWIR, shape and scale are passed to the system for Weibull distributions instead of average and standard deviation.

Parameter uncertainty information is provided in Appendixes 8D, 8E, and 8F. This information includes maximum likelihood estimates for location and scale parameters for each of the 3 two-parameter models (Appendix 8D), as well as estimates of the standard errors for each

Table 8-2. Chi-Square Statistics and *p*-Values for Goodness-of-Fit Tests for Models Fit by Maximum Likelihood

Parameter	Age Cohort	N	M	First	Second	Third	CHISQ GAM	CHISQ LOG	CHISQ WEI	PGOF GAM	PGOF LOG	PGOF WEI
beef	6-11	38	7	lognormal	gamma	Weibull	12.77	4.86	15.01	0.047	0.562	0.02
beef	12-19	41	7	gamma	Weibull	lognormal	4.06	5.07	4.81	0.668	0.534	0.569
beef	farmer	182	9	lognormal	gamma	Weibull	35.28	5.57	64.54	0	0.695	0
bodywt	<1	356	9	gamma	lognormal	Weibull	20.21	22.23	37.05	0.01	0.005	0
bodywt	1-5	3,762	9	lognormal	gamma	Weibull	87.53	52.03	706.3	0	0	0
bodywt	6-11	1,725	9	lognormal	gamma	Weibull	156.9	105.1	585.1	0	0	0
bodywt	12-19	2,615	9	lognormal	gamma	Weibull	113.8	70.32	625.2	0	0	0
bodywt	20+	12,504	9	lognormal	gamma	Weibull	420.7	215	2584	0	0	0
cumTroom	all ages	6,661	8	lognormal	gamma	Weibull	626	426.6	1067	0	0	0
drinkH_2O	<1	403	5	Weibull	gamma	lognormal	14.62	44.72	12.59	0.006	0	0.013
drinkH_2O	1-5	3,200	9	gamma	Weibull	lognormal	60.84	290.5	113.3	0	0	0
drinkH_2O	6-11	2,405	9	gamma	Weibull	lognormal	38.77	243	55.76	0	0	0
drinkH_2O	12-19	5,801	9	gamma	Weibull	lognormal	93.37	625.2	131.8	0	0	0
drinkH_2O	20+	13,394	9	gamma	Weibull	lognormal	172.2	624.3	467	0	0	0
expfruit	1-5	49	6	gamma	Weibull	lognormal	6.38	8.38	6.79	0.271	0.137	0.236
expfruit	6-11	68	7	lognormal	Weibull	gamma	15.83	6.84	15.25	0.015	0.336	0.018
expfruit	12-19	50	7	lognormal	Weibull	gamma	16.71	14.11	16.52	0.01	0.028	0.011
expfruit	farmer	112	9	lognormal	gamma	Weibull	13.43	7.7	17.11	0.098	0.463	0.029
expfruit	home gard.	596	9	lognormal	gamma	Weibull	70.31	14.52	93.27	0	0.069	0
expveg	1-5	105	7	gamma	Weibull	lognormal	11.33	16.23	11.44	0.079	0.013	0.076
expveg	6-11	134	8	lognormal	Weibull	gamma	11.9	8.26	9.2	0.104	0.31	0.238
expveg	12-19	143	8	gamma	Weibull	lognormal	16.09	35.88	16.11	0.024	0	0.024
expveg	farmer	207	8	lognormal	gamma	Weibull	19.03	18.92	20.18	0.008	0.008	0.005
expveg	home gard.	1,361	9	Weibull	gamma	lognormal	81.63	98.94	77.22	0	0	0
fish	all ages	1,053	5	lognormal	Weibull	gamma	41.9	6.74	20.55	0	0.15	0
milk	<1	20	6	Weibull	lognormal		a	22.51	8.63	a	0	0.125
milk	1-5	40	7	Weibull	gamma	lognormal	3.25	9.1	1.72	0.777	0.168	0.943
milk	6-11	20	7	Weibull	gamma	lognormal	0.71	3.08	0.25	0.994	0.798	1
milk	12-19	20	7	Weibull	gamma	lognormal	1.53	6.66	1.17	0.958	0.354	0.978
milk	farmer	63	7	Weibull	gamma	lognormal	14.29	26.01	13.43	0.027	0	0.037
profruit	12-19	20	7	lognormal	Weibull	gamma	4.58	1.74	4.21	0.599	0.942	0.649
profruit	20+	106	7	lognormal	Weibull	gamma	19.94	8.93	17.95	0.003	0.177	0.006
profruit	all ages	173	9	lognormal	Weibull	gamma	68.45	12.82	65.92	0	0.118	0
profruit	home gard.	146	9	lognormal	Weibull	gamma	38.27	14.05	36.12	0	0.08	0
proveg	1-5	53	7	lognormal	gamma	Weibull	15.51	11.09	19.95	0.017	0.086	0.003
proveg	6-11	63	7	lognormal	gamma	Weibull	8.63	6.35	11.91	0.196	0.385	0.064
proveg	12-19	51	7	lognormal	gamma	Weibull	13.29	6.71	17.2	0.039	0.349	0.009
proveg	farmer	142	9	lognormal	gamma	Weibull	83.34	13.04	108.5	0	0.11	0
proveg	home gard.	602	9	lognormal	gamma	Weibull	181.4	17.93	318.4	0	0.022	0
rootveg	1-5	45	7	lognormal	Weibull	gamma	8.79	4.22	7.72	0.186	0.647	0.259
rootveg	6-11	67	7	Weibull	gamma	lognormal	4.68	10.45	3.83	0.586	0.107	0.699
rootveg	12-19	76	7	Weibull	lognormal		a	23.43	6.82	a	0.001	0.337
rootveg	farmer	136	9	lognormal	gamma	Weibull	43.62	11.16	59.84	0	0.193	0
rootveg	home gard.	682	9	Weibull	gamma	lognormal	27.9	102	26.91	0	0	0.001
showerT	all ages	3,547	10	gamma	lognormal	Weibull	485.2	524.4	935.2	0	0	0

a Information on gamma distribution is not available because the nonlinear optimization routine (the SAS NLIN procedure) used for model fitting would not converge on a solution.

CHISQ = chi-square; GAM = gamma; LOG = lognormal; PGOF = *p*-values for the goodness of fit; WEI = Weibull.

of these parameter estimates (Appendixes 8E and 8F) and estimates of the correlation between the location and scale parameters (Appendix 8F). Model parameterizations based on location and scale parameters are provided in Appendix 8G.

To generate the required bivariate normal random variables, the estimated variances and the estimated correlations between the location and scale parameters must be known. The estimated correlations are in Appendix 8F. The estimated variances are the squares of the standard errors in Appendixes 8E and 8F, which contain two sets of estimated standard errors. The first set, generated using the assumed asymptotic normality of the maximum likelihood estimate as in Burmaster and Thompson (1998), can be used when sample sizes are available. The second employs a nonlinear regression method (Jennrich and Ralston, 1979) that uses the number of percentiles but not the sample size in calculating statistics; it can be used when sample sizes are not available.

The regression method infers the precision of the estimates solely from the fit to the data (i.e., from the regression mean squared error). It has two advantages: it produces statistics when the sample size is unknown, and it absorbs model lack of fit in precision estimates. Therefore, the regression method takes both model uncertainty and parameter uncertainty into account. Accordingly, the standard errors from the regression method tend to be much larger than those based on the asymptotic normality of the maximum likelihood estimate. Although the regression method has been cited as more valid and recommended for use in all cases (Huber, 1967; White, 1982), some users may opt for standard errors based on the asymptotic normality of the maximum likelihood estimate when these are available, reserving the regression statistics only for situations with unknown sample sizes.

8.3.1.4 <u>Selection of HWIR Distributions and Measures of Uncertainty</u>. For each variable listed in Table 8-2, the distribution with the first chi-square rank was selected for use in the HWIR analysis. For all selected distribution types except Weibull, the arithmetic means and standard deviations in Appendix 8B were passed to the HWIR modeling system; for Weibull, shape (α) and scale (β) parameters from Appendix 8C were used.[3] These statistics also are provided by exposure factor variable in Section 8.4.

Parameters that had to be analyzed separately because of lack of adequate percentile data include beef consumption for those aged 1 to 5 years; protected fruit consumption for those aged 1 to 5 years, those aged 6 to 11 years, and farmers; breast milk consumption; soil ingestion; and inhalation rate. For these parameters, either distributions for similar parameters were assumed to apply or a default lognormal model was selected. Methods and results for these parameters also are provided in Section 8.4.

- # Percentile data were not available for **beef consumption** for 1- to 5-year-olds. In this case, the lognormal model was used because, among the other age groups, it was the best-fitted model in all but one case.

[3] The implied Weibull cumulative distribution function (CDF) is, therefore, 1-exp ($-\alpha * x^{\beta}$).

Percentile data were not provided for **consumption of protected fruit** (child1, child2, and farmer). Therefore, the lognormal model was considered the most appropriate because lognormal fits the best in other age groups for protected fruit and protected vegetables.

The lognormal model was assumed for **breast milk consumption** (those aged 1 to 12 months) because no other applicable data were available. In the EFH data, the population mean for breast milk consumption was 688 mL/d and the population standard deviation (for CV=1.5) was 1,032 mL/d, both of which were used for this analysis.

Similarly, the lognormal model was used for **soil consumption** for all age groups because of limited percentile data. Parameter estimates were obtained by assuming CV=0.5, 1, and 1.5 (see Section 8.4.12).

No percentile data were available for **inhalation rate**, and the default lognormal model was assumed. An analysis of inhalation data by Myers et al. (1998) found that for those younger than age 3, CV was close to 70 percent; for other age groups, it was close to 30 percent. The lognormal distribution was fitted by using CV=70 percent for the child1 age group; CV=50 percent [(30+70)/2] for the child2 age group; and CV=30 percent for the child3, child4, and adult age groups (see Section 8.4.13).

8.3.2 Fixed Parameters

Certain parameters were fixed nationally, based on central tendency values from the best available source (usually EFH recommendations), either because no regional variability was expected or because the available data were not adequate to generate national distributions. These fixed parameters include exposure frequency; shower frequency; and fraction contaminated for drinking water, soil, exposed vegetables, root vegetables, protected vegetables, exposed fruit, protected fruit, and fish. Fraction-contaminated values indicate the amount of foodstuff grown or produced on a farm or in the garden in question and correspond to the fraction homegrown or home-produced from the EFH (Table 13-71). Exposure duration was fixed based on EPA policy. Human lifetime was set at a projected average value from the EFH (Table 8-1, U.S. EPA, 1997a). Section 8.4.15 provides additional details on the fixed HWIR exposure factors.

8.3.3 Quality Assurance/Quality Control

The following QA/QC measures were used to minimize errors associated with data collection, analysis, and processing:

Raw Data Extraction and Entry–One hundred percent checks were made of all exposure factor data extracted from the EFH and entered into spreadsheets for statistical analysis.

- # **Statistical Analysis**–All statistical analyses were conducted in duplicate using replicate data inputs to ensure reproducibility of results.

- # **Data Entry for the HWIR Modeling System–**Data entry was checked 100 percent against the statistical analysis outputs. All conversions performed by the data processing system were checked against hand calculations.

- # **Documentation–**All hardcopy data sources are organized and maintained in a formal filing system, along with documentation of QA/QC procedures.

Detailed QA/QC documentation can be made available upon request.

8.4 Results by Exposure Factor

Table 8-3 summarizes the exposure factor stochastic or distributed input data used for the HWIR analysis. The following sections describe how these data were collected and processed by exposure factor. Each section includes a table showing the raw EFH data used to develop the distributions, along with the final distributional statistics. Fixed factors are discussed in Section 8.4.15.

8.4.1 Body Weight

EFH Data–Body Weight (kg)													HWIR Distributions		
Age Cohort	N	Data Mean	Data SDev	P05	P10	P15	P25	P50	P75	P85	P90	P95	Distribution	Pop-Estd Mean	Pop-Estd SDev
<1	356	9.102	1.287	7.053	7.451	7.852	8.252	9.151	9.752	10.4	10.65	11.15	gamma	9.09	1.23
1-5	3,762	15.52	3.719	12.5	13.1	13.45	14.03	15.26	16.67	17.58	18.32	19.45	lognormal	15.5	2.05
6-11	1,725	30.84	9.561	22.79	24.05	25.07	26.44	29.58	33.44	36.82	39.66	43.5	lognormal	30.7	5.96
12-19	2,615	58.45	13.64	43.84	46.52	48.31	50.94	56.77	63.57	68.09	71.98	79.52	lognormal	58.2	10.2
20+	12,504	71.41	15.45	52.86	55.98	58.21	61.69	69.26	78.49	84.92	89.75	97.64	lognormal	71.2	13.3

N = number of samples; P05-P95 = percentiles; Pop-Estd = population-estimated; SDev = standard deviation.

Body weight data were obtained from Tables 7-2, 7-3, 7-4, 7-5, 7-6, and 7-7 of the EFH (U.S. EPA, 1997a). Data (in kg) were presented by age and gender. Weighted averages of percentiles, means, and standard deviations were calculated for child1 (<1 year old), child2 (1 to 5 years old), child3 (6 to 11 years old), child4 (12 to 19 years old), and adult age groups; male and female data were weighted and combined for each age group. These percentile data were used as the basis of fitting distributions. These data were analyzed to fit parametric models (gamma, lognormal, and Weibull) using maximum likelihood estimation. Measures of goodness of fit were used to select the most appropriate model.

Table 8-3. Summary of HWIR Human Exposure Factor Data Used in Modeling: Stochastic Variables

Code	Description	Units	Distribution Type	Mean	Source	SDev	Source	Min	Source	Max	Source
Bri_cr1	inhalation (breathing) rate (child1 resident)	m^3/d	lognormal	4.5	LOG mean	3.15	LOG SDEV	0.5	EFH (newborn)	30	2*(mean+3SD)
Bri_cr2	inhalation (breathing) rate (child2 resident)	m^3/d	lognormal	7.55	LOG mean	3.78	LOG SDEV	1	0.5*(mean-3SD)	40	2*(mean+3SD)
Bri_cr3	inhalation (breathing) rate (child3 resident)	m^3/d	lognormal	11.75	LOG mean	3.53	LOG SDEV	1	0.5*(mean-3SD)	45	2*(mean+3SD)
Bri_cr4	inhalation (breathing) rate (child4 resident)	m^3/d	lognormal	14.0	LOG mean	4.2	LOG SDEV	1	0.5*(mean-3SD)	55	2*(mean+3SD)
Bri_r	inhalation (breathing) rate (adult resident)	m^3/d	lognormal	13.3	LOG mean	3.99	LOG SDEV	1	0.5*(mean-3SD)	50	2*(mean+3SD)
BWa	body weight (adult)	kg	lognormal	71.2	LOG mean	13.3	LOG SDEV	15	0.5*(mean-3SD)	300	Prof. Judgement
BWc1	body weight (child1)	kg	gamma	9.09	GAM mean	1.23	GAM SDEV	2	0.5*(mean-3SD)	26	2*(mean+3SD)
BWc2	body weight (child2)	kg	lognormal	15.5	LOG mean	2.05	LOG SDEV	4	0.5*(mean-3SD)	50	Prof. Judgment
BWc3	body weight (child3)	kg	lognormal	30.7	LOG mean	5.96	LOG SDEV	6	0.5*(mean-3SD)	200	Prof. Judgment
BWc4	body weight (child4)	kg	lognormal	58.2	LOG mean	10.2	LOG SDEV	13	0.5*(mean-3SD)	300	Prof. Judgment
CRb_af	consumption rate: beef (adult farmer)	g WW/kg/d	lognormal	2.5	LOG mean	2.69	LOG SDEV	0		23	2*(P99)
CRb_cf_2	consumption rate: beef (child2 farmer)	g WW/kg/d	lognormal	3.88	LOG mean	4.71	LOG SDEV	0		36	2*(mean+3SD)
CRb_cf_3	consumption rate: beef (child3 farmer)	g WW/kg/d	lognormal	3.88	LOG mean	4.71	LOG SDEV	0		36	2*(mean+3SD)
CRb_cf_4	consumption rate: beef (child4 farmer)	g WW/kg/d	gamma	1.77	GAM mean	1.12	GAM SDEV	0		10	2*(mean+3SD)
CRbm_cr_1	consumption rate: breast milk (child1)	mL/d	lognormal	688	EFH data	1,032	EFH data	0		1,200	EFH max. = 1165
Crfr_cf_2	consumption rate: exposed fruit (child2 farmer)	g WW/kg/d	gamma	2.25	GAM mean	1.89	GAM SDEV	0		16	2*(mean+3SD)
CRfr_cf_3	consumption rate: exposed fruit (child3 farmer)	g WW/kg/d	lognormal	2.78	LOG mean	5.12	LOG SDEV	0		36	2*(mean+3SD)
CRfr_cf_4	consumption rate: exposed fruit (child4 farmer)	g WW/kg/d	lognormal	1.54	LOG mean	2.44	LOG SDEV	0		18	2*(mean+3SD)
CRfr_cg_2	consumption rate: exposed fruit (child2 gardener)	g WW/kg/d	gamma	2.25	GAM mean	1.89	GAM SDEV	0		16	2*(mean+3SD)
CRfr_cg_3	consumption rate: exposed fruit (child3 gardener)	g WW/kg/d	lognormal	2.78	LOG mean	5.12	LOG SDEV	0		36	2*(mean+3SD)
CRfr_cg_4	consumption rate: exposed fruit (child4 gardener)	g WW/kg/d	lognormal	1.54	LOG mean	2.44	LOG SDEV	0		18	2*(mean+3SD)

(continued)

Table 8-3. (continued)

Code	Description	Units	Distribution Type	Mean	Source	SDev	Source	Min	Source	Max	Source
CRfr_f	consumption rate: exposed fruit (farmer)	g WW/kg/d	lognormal	2.36	LOG mean	3.33	LOG SDEV	0		31	2*(P99)
CRfr_g	consumption rate: exposed fruit (gardener)	g WW/kg/d	lognormal	1.57	LOG mean	2.30	LOG SDEV	0		26	2*(P99)
CRfs_a	consumption rate: fish (adult)	g/d	lognormal	6.48	LOG mean	19.9	LOG SDEV	0		1,500	EFH-subsist
CRfs_c_2	consumption rate: fish (child2)	g/d	lognormal	6.48	LOG mean	19.9	LOG SDEV	0		1,500	EFH-subsist
CRfs_c_3	consumption rate: fish (child3)	g/d	lognormal	6.48	LOG mean	9.9	LOG SDEV	0		1,500	EFH-subsist
CRfs_c_4	consumption rate: fish (child4)	g/d	lognormal	6.48	LOG mean	19.9	LOG SDEV	0		1,500	EFH-subsist
CRl_cf_2	consumption rate: exposed vegetables (child2 farmer)	g WW/kg/d	gamma	2.55	GAM mean	2.58	GAM SDEV	0		21	2*(mean+3SD)
CRl_cf_3	consumption rate: exposed vegetables (child3 farmer)	g WW/kg/d	lognormal	1.64	LOG mean	3.95	LOG SDEV	0		27	2*(mean+3SD)
CRl_cf_4	consumption rate: exposed vegetables (child4 farmer)	g WW/kg/d	gamma	1.08	GAM mean	1.13	GAM SDEV	0		11	2*(P99)
CRl_cg2	consumption rate: exposed vegetables (child2 gardener)	g WW/kg/d	gamma	2.55	GAM mean	2.58	GAM SDEV	0		21	2*(mean+3SD)
CRl_cg3	consumption rate: exposed vegetables (child3 gardener)	g WW/kg/d	lognormal	1.64	LOG mean	3.95	LOG SDEV	0		27	2*(mean+3SD)
CRl_cg4	consumption rate: exposed vegetables (child4 gardener)	g WW/kg/d	gamma	1.08	GAM mean	1.13	GAM SDEV	0		11	2*(P99)
CRl_f	consumption rate: exposed vegetables (adult farmer)	g WW/kg/d	lognormal	2.38	LOG mean	3.5	LOG SDEV	0		26	2*(mean+3SD)
CRl_g	consumption rate: exposed vegetables (gardener)	g WW/kg/d	Weibull	0.89	WEI shape	1.48	WEI scale	0		21	2*(P99)
CRm_af	consumption rate: milk (adult farmer)	g WW/kg/d	Weibull	1.25	WEI shape	17.45	WEI scale	0		111	2*(mean+3SD)
CRm_cf_2	consumption rate: milk (child2 farmer)	g WW/kg/d	Weibull	1.7	WEI shape	26.47	WEI scale	0		133	2*(mean+3SD)
CRm_cf_3	consumption rate: milk (child3 farmer)	g WW/kg/d	Weibull	1.56	WEI shape	14.82	WEI scale	0		79	2*(mean+3SD)
CRm_cf_4	consumption rate: milk (child4 farmer)	g WW/kg/d	Weibull	1.14	WEI shape	6.52	WEI scale	0		45	2*(mean+3SD)
CRpfr_cf_2	consumption rate: protected fruit (child2 farmer)	g WW/kg/d	lognormal	6.5	LOG mean	15.9	LOG SDEV	0		108	2*(mean+3SD)
CRpfr_cf_3	consumption rate: protected fruit (child3 farmer)	g WW/kg/d	lognormal	6.5	LOG mean	15.9	LOG SDEV	0		108	2*(mean+3SD)
CRpfr_cf_4	consumption rate: protected fruit (child4 farmer)	g WW/kg/d	lognormal	2.91	LOG mean	6.39	LOG SDEV	0		44	2*(mean+3SD)

(continued)

Table 8-3. (continued)

Code	Description	Units	Distribution Type	Mean	Source	SDev	Source	Min	Source	Max	Source
CRpfr_cg_2	consumption rate: protected fruit (child2 gardener)	g WW/kg/d	lognormal	6.5	LOG mean	15.9	LOG SDEV	0		108	2*(mean+3SD)
CRpfr_cg_3	consumption rate: protected fruit (child3 gardener)	g WW/kg/d	lognormal	6.5	LOG mean	15.9	LOG SDEV	0		108	2*(mean+3SD)
CRpfr_cg_4	consumption rate: protected fruit (child4 gardener)	g WW/kg/d	lognormal	2.91	LOG mean	6.39	LOG SDEV	0		44	2*(mean+3SD)
CRpfr_f	consumption rate: protected fruit (adult farmer)	g WW/kg/d	lognormal	6.67	LOG mean	17.7	LOG SDEV	0		120	2*(mean+3SD)
CRpfr_g	consumption rate: protected fruit (adult gardener)	g WW/kg/d	lognormal	6.63	LOG mean	15.7	LOG SDEV	0		108	2*(mean+3SD)
CRpl_cf_2	consumption rate: protected vegetables (child2 farmer)	g WW/kg/d	lognormal	1.88	LOG mean	1.98	LOG SDEV	0		16	2*(mean+3SD)
CRpl_cf_3	consumption rate: protected vegetables (child3 farmer)	g WW/kg/d	lognormal	1.07	LOG mean	1.04	LOG SDEV	0		8	2*(mean+3SD)
CRpl_cf_4	consumption rate: protected vegetables (child4 farmer)	g WW/kg/d	lognormal	0.77	LOG mean	0.69	LOG SDEV	0		6	2*(mean+3SD)
CRpl_cg_2	consumption rate: protected vegetables (child2 gardener)	g WW/kg/d	lognormal	1.88	LOG mean	1.98	LOG SDEV	0		16	2*(mean+3SD)
CRpl_cg_3	consumption rate: protected vegetables (child3 gardener)	g WW/kg/d	lognormal	1.07	LOG mean	1.04	LOG SDEV	0		8	2*(mean+3SD)
CRpl_cg_4	consumption rate: protected vegetables (child4 gardener)	g WW/kg/d	lognormal	0.77	LOG mean	0.69	LOG SDEV	0		6	2*(mean+3SD)
CRpl_f	consumption rate: protected vegetables (adult farmer)	g WW/kg/d	lognormal	1.27	LOG mean	1.85	LOG SDEV	0		18	2*(P99)
CRpl_g	consumption rate: protected vegetables (adult gardener)	g WW/kg/d	lognormal	1.01	LOG mean	1.19	LOG SDEV	0		13	2*(P99)
CRr_cf_2	consumption rate: root vegetables (child2 farmer)	g WW/kg/d	lognormal	2.31	LOG mean	6.05	LOG SDEV	0		41	2*(mean+3SD)
CRr_cf_3	consumption rate: root vegetables (child3 farmer)	g WW/kg/d	Weibull	0.68	WEI shape	1.06	WEI scale	0		15	2*(mean+3SD)
CRr_cf_4	consumption rate: root vegetables (child4 farmer)	g WW/kg/d	Weibull	0.84	WEI shape	0.91	WEI scale	0		9	2*(mean+3SD)
CRr_cg_2	consumption rate: root vegetables (child2 gardener)	g WW/kg/d	lognormal	2.31	LOG mean	6.05	LOG SDEV	0		41	2*(mean+3SD)
CRr_cg_3	consumption rate: root vegetables (child3 gardener)	g WW/kg/d	Weibull	0.68	WEI shape	1.06	WEI scale	0		15	2*(mean+3SD)
CRr_cg_4	consumption rate: root vegetables (child4 gardener)	g WW/kg/d	Weibull	0.84	WEI shape	0.91	WEI scale	0		9	2*(mean+3SD)
CRr_f	consumption rate: root vegetables (farmer)	g WW/kg/d	lognormal	1.45	LOG mean	2.06	LOG SDEV	0		15	2*(mean+3SD)
CRr_g	consumption rate: root vegetables (gardener)	g WW/kg/d	Weibull	0.87	WEI shape	1.07	WEI scale	0		15	2*(P99)

(continued)

Table 8-3. (continued)

Code	Description	Units	Distribution Type	Mean	Source	SDev	Source	Min	Source	Max	Source
CRs_cr2	ingestion rate: soil (child2 resident)	kg/d	lognormal	0.0001	LOG mean	1.5E-04	LOG SDEV	5E-07	Prof. Judgment	0.03	EFH-P75 pica
CRs_cr3	ingestion rate: soil (child3 resident)	kg/d	lognormal	5E-05	LOG mean	7.5E-05	LOG SDEV	5E-07	Prof. Judgment	0.03	EFH-P75 pica
CRS_cr4	ingestion rate: soil (child4 resident)	kg/d	lognormal	5E-05	LOG mean	7.5E-05	LOG SDEV	5E-07	Prof. Judgment	0.03	EFH-P75 pica
CRs_r	ingestion rate: soil (adult resident)	kg/d	lognormal	5E-05	LOG mean	7.5E-05	LOG SDEV	5E-07	Prof. Judgment	0.03	EFH-P75 pica
CRw_cr1	ingestion rate: drinking water (child1 resident)	mL/d	Weibull	1.16	WEI shape	318.60	WEI scale	0	Prof. Judgment	2,200	2*(mean+3SD)
CRw_cr2	ingestion rate: drinking water (child2 resident)	mL/d	gamma	698	GAM mean	406	GAM SDEV	26	0.5*(P01)	3,840	2*(P99)
CRw_cr3	ingestion rate: drinking water (child3 resident)	mL/d	gamma	787	GAM mean	430	GAM SDEV	34	0.5*(P01)	4,200	2*(mean+3SD)
CRw_cr4	ingestion rate: drinking water (child4 resident)	mL/d	gamma	965	GAM mean	574	GAM SDEV	33	0.5*(P01)	5,400	2*(P99)
CRw_r	ingestion rate: drinking water (adult resident)	mL/d	gamma	1,383	GAM mean	703	GAM SDEV	104	0.5*(P01)	11,000	EFH-active, hot
t_sb	time in shower and bathroom	min	lognormal	34	LOG mean	34.6	LOG SDEV	1	Prof. Judgment	240	Prof. Judgment
t_shower	shower time	min	gamma	16.7	GAM mean	9.91	GAM SDEV	1	Prof. Judgment	60	Prof. Judgment

8.4.2 Drinking Water Intake

EFH Data–Drinking Water Intake (mL/d)													HWIR Distributions		
Age Cohort	**N**	**Data Mean**	**Data SDev**	**P01**	**P05**	**P10**	**P25**	**P50**	**P75**	**P90**	**P95**	**P99**	**Distribution**	**Pop-Estd Mean**	**Pop-Estd SDev**
< 1	403	302.7	258.5				100.3	255.4	413.2	666.3	780.3		Weibull	302	261
1-5	3,200	697.1	401.5	51.62	187.6	273.5	419.2	616.5	900.8	1,236	1,473	1,917	gamma	698	406
6-11	2,405	787	417	68	241	318	484	731	1,016	1,338	1,556	1,998	gamma	787	430
12-19	5,801	963.2	560.6	65.15	241.4	353.8	574.4	868.5	1,247	1,694	2,033	2,693	gamma	965	574
20+	13,394	1,384	721.6	207.6	457.5	607.3	899.6	1,275	1,741	2,260	2,682	3,737	gamma	1,383	703

N = number of samples; P01-P99 = percentiles; Pop-Estd = population-estimated; SDev = standard deviation.

Drinking water intake data were obtained from Table 3-6 of the EFH (U.S. EPA, 1997a). Data (in mL/d) were presented by age groups. Weighted averages of percentiles, means, and standard deviations were calculated for child1, child2, child3, child4, and adult age groups. Percentile data were used to fit parametric models (gamma, lognormal, and Weibull) using maximum likelihood estimation. Measures of goodness of fit were used to select the most appropriate model.

8.4.3 Exposed Fruit Consumption

EFH Data–Exposed Fruit Consumption (g WW/kg/d)													HWIR Distributions		
Age Cohort	**N**	**Data Mean**	**Data SDev**	**P01**	**P05**	**P10**	**P25**	**P50**	**P75**	**P90**	**P95**	**P99**	**Distribution**	**Pop-Estd Mean**	**Pop-Estd SDev**
1-5	49	2.6	3.947			0.373	1	1.82	2.64	5.41	6.07		gamma	2.25	1.89
6-11	68	2.52	3.496		0.171	0.373	0.619	1.11	2.91	6.98	11.7		lognormal	2.78	5.12
12-19	50	1.33	1.457		0.123	0.258	0.404	0.609	2.27	3.41	4.78		lognormal	1.54	2.44
farmer	112	2.32	2.646	0.072	0.276	0.371	0.681	1.3	3.14	5	6.12	15.7	lognormal	2.36	3.33
home gard.	596	1.55	2.226	0.042	0.158	0.258	0.449	0.878	1.73	3.41	5	12.9	lognormal	1.57	2.3

N = number of samples; P01-P99 = percentiles; Pop-Estd = population-estimated; SDev = standard deviation.

Data for consumption of homegrown exposed fruit were obtained from Table 13-61 of the EFH (U.S. EPA, 1997b). Data (in g WW/kg/d) were presented by age groups and for farmers and home gardeners (adults). For the child2 age group, data were only available for those ages 3 to 5 years (not available for 1- to 2-year-olds); therefore, these data were used for the entire 1- to 5-year-old age group (child2). Percentile data were used to fit parametric models (gamma, lognormal, and Weibull) using maximum likelihood estimation. Measures of goodness of fit were used to select the most appropriate model. The fraction of exposed fruit intake that is home-produced is 0.116 for households that garden and 0.328 for households that farm (Table 13-71, U.S. EPA, 1997b).

8.4.4 Protected Fruit Consumption

EFH Data–Protected Fruit Consumption (g WW/kg/d)													HWIR Distributions		
Age Cohort	**N**	**Data Mean**	**Data SDev**	**P01**	**P05**	**P10**	**P25**	**P50**	**P75**	**P90**	**P95**	**P99**	**Distribution**	**Pop-Estd Mean**	**Pop-Estd SDev**
1-5		ND	ND										lognormal	6.5	15.9
6-11		ND	ND										lognormal	6.5	15.9
12-19	20	2.96	4.441		0.16	0.283	0.393	1.23	2.84	7.44	11.4		lognormal	2.91	6.39
20+	106	5.338	7.174		0.276	0.342	0.82	2.127	8.022	15.25	19.8		lognormal	6.67	17.7
all ages	173	5.74	8.221	0.15	0.266	0.335	0.933	2.34	7.45	16	19.7	47.3	lognormal	6.5	15.9
farmer		ND	ND										lognormal	6.67	17.7
home gard.	146	5.9	8.422	0.117	0.265	0.335	1.16	2.42	7.46	16	19.1	47.3	lognormal	6.63	15.7

N = number of samples; P01-P99 = percentiles; Pop-Estd = population-estimated; SDev = standard deviation.

Data for consumption of homegrown protected fruit were obtained from Table 13-62 of the EFH (U.S. EPA, 1997b). Data (in g WW/kg/d) were presented for those 12 to 19 years, 20 to 39 years, 40 to 69 years, all ages, and home gardeners. Available percentile data were used to fit parametric models (gamma, lognormal, and Weibull) using maximum likelihood estimation. Measures of goodness of fit were used to select the most appropriate model.

Data were not available for farmers or those aged 1 to 2, 3 to 5, and 6 to 11 years. For the child2 and child3 age groups, the lognormal model is most appropriate because lognormal fits the best in other age groups for protected fruit and vegetables; the population estimated mean and standard deviation for all age groups were used for the analysis (normalized to body weight). For farmers, the population estimated mean and standard deviation for those older than 20 years (derived from the weighted average of means and standard deviations of those ages 20 to 39 years and those ages 40 to 69 years) were used for the analysis; lognormal also fits the percentile data best for those older than 20 years. The fraction of protected fruit intake that is home-produced is 0.094 for households that garden and 0.03 for households that farm (Table 13-71, U.S. EPA, 1997b).

8.4.5 Exposed Vegetable Consumption

EFH Data–Exposed Vegetable Consumption (g WW/kg/d)													HWIR Distributions		
Age Cohort	**N**	**Data Mean**	**Data SDev**	**P01**	**P05**	**P10**	**P25**	**P50**	**P75**	**P90**	**P95**	**P99**	**Distribution**	**Pop-Estd Mean**	**Pop-Estd SDev**
1-5	105	2.453	2.675		0.102	0.37	0.833	1.459	3.226	6.431	8.587		gamma	2.55	2.58
6-11	134	1.39	2.037		0.044	0.094	0.312	0.643	1.6	3.22	5.47	13.3	lognormal	1.64	3.95
12-19	143	1.07	1.128		0.029	0.142	0.304	0.656	1.46	2.35	3.78	5.67	gamma	1.08	1.13h
farmer	207	2.17	2.316		0.184	0.372	0.647	1.38	2.81	6.01	6.83	10.3	lognormal	2.38	3.5
home gard.	1,361	1.57	2.029	0.003	0.089	0.168	0.413	0.889	1.97	3.63	5.45	10.3	Weibull	1.57	1.76

N = number of samples; P01-P99 = percentiles; Pop-Estd = population-estimated; SDev = standard deviation.

Data for consumption of homegrown exposed vegetables were obtained from Table 13-63 of the EFH (U.S. EPA, 1997b). Data (in g WW/kg/d) were presented for those ages 1 to 2, 3 to 5, 6 to 11, 12 to 19, 20 to 39, and 40 to 69 years, as well as farmers and home gardeners. Weighted averages of percentiles, means, and standard deviations were calculated for the child2 age group (combining groups of those ages 1 to 2 years and 3 to 5 years). Percentile data were used to fit parametric models (gamma, lognormal, and Weibull) using maximum likelihood estimation. Measures of goodness of fit were used to select the most appropriate model.

8.4.6 Root Vegetable Consumption

EFH Data–Root Vegetable Consumption (g WW/kg/d)													HWIR Distributions		
Age Cohort	N	Data Mean	Data SDev	P01	P05	P10	P25	P50	P75	P90	P95	P99	Distribution	Pop-Estd Mean	Pop-Estd SDev
1-5	45	1.886	2.371		0.081	0.167	0.291	0.686	2.653	5.722	7.502		lognormal	2.31	6.05
6-11	67	1.32	1.752		0.014	0.036	0.232	0.523	1.63	3.83	5.59		Weibull	1.38	2.07
12-19	76	0.937	1.037		0.008	0.068	0.269	0.565	1.37	2.26	3.32		Weibull	0.99	1.19
farmer	136	1.39	1.469	0.111	0.158	0.184	0.365	0.883	1.85	3.11	4.58	7.47	lognormal	1.45	2.06
home gard.	682	1.15	1.494	0.005	0.036	0.117	0.258	0.674	1.5	2.81	3.64	7.47	Weibull	1.15	1.32

N = number of samples; P01-P99 = percentiles; Pop-Estd = population-estimated; SDev = standard deviation.

Homegrown root vegetable consumption data were obtained from Table 13-65 of the EFH (U.S. EPA, 1997b). Data (in g WW/kg/d) were presented for those ages 1 to 2, 3 to 5, 6 to 11, 12 to 19, 20 to 39, and 40 to 69 years, as well as farmers and home gardeners. Weighted averages of percentiles, means, and standard deviations were calculated for the child2 age group (combining groups of those ages 1 to 2 and 3 to 5 years). Percentile data were used to fit parametric models (gamma, lognormal, and Weibull) using maximum likelihood estimation. Measures of goodness of fit were used to select the most appropriate model.

8.4.7 Protected Vegetable Consumption

EFH Data–Protected Vegetable Consumption (g WW/kg/d)													HWIR Distributions		
Age Cohort	N	Data Mean	Data SDev	P01	P05	P10	P25	P50	P75	P90	P95	P99	Distribution	Pop-Estd Mean	Pop-Estd SDev
1-5	53	1.76	1.79		0.265	0.408	0.829	1.397	2.066	3.053	6.812		lognormal	1.88	1.98
6-11	63	1.1	1.064		0.208	0.318	0.387	0.791	1.31	2.14	3.12		lognormal	1.07	1.04
12-19	51	0.776	0.622		0.161	0.239	0.354	0.583	0.824	1.85	2.2		lognormal	0.77	0.69
farmer	142	1.3	1.728	0.087	0.166	0.209	0.337	0.599	1.4	3.55	5.4	9.23	lognormal	1.27	1.85
home gard.	602	1.01	1.161	0.103	0.153	0.192	0.336	0.642	1.21	2.32	3.05	6.49	lognormal	1.01	1.19

N = number of samples; P01-P99 = percentiles; Pop-Estd = population-estimated; SDev = standard deviation.

Homegrown protected vegetable consumption data were obtained from Table 13-64 of the EFH (U.S. EPA, 1997b). Data (in g WW/kg/d) were presented for those ages 1 to 2, 3 to 5, 6 to 11, 12 to 19, 20 to 39, and 40 to 69 years, as well as farmers and home gardeners. Weighted averages of percentiles, means, and standard deviations were calculated for the child2 age group (combining groups of those ages 1 to 2 and 3 to 5 years). Percentile data were used to fit parametric models (gamma, lognormal, and Weibull) using maximum likelihood estimation. Measures of goodness of fit were used to select the most appropriate model.

8.4.8 Dairy Products (Milk) Consumption

EFH Data–Milk Consumption (g WW/kg/d)											HWIR Distributions		
Age Cohort	N	Data Mean	Data SDev	P05	P10	P25	P50	P75	P90	P95	Distribution	Pop-Estd Mean	Pop-Estd SDev
<1	1	62.74	94.1		0.61	24.68	45.78	91.12	136.7	170.9	Weibull	65.4	78.7
1-5	2	23.71	35.86	2.98	7.47	13.56	21.5	32.22	42.63	49.62	Weibull	23.6	14.3
6-11	1	13.33	20	1.81	3.54	6.72	11.88	18.58	25.38	28.76	Weibull	13.3	8.7
12-19	1	6.293	9.44	0.27	0.61	2.31	5.29	9.2	12.75	15.12	Weibull	6.23	5.49
farmer	63	17.1	15.8	0.736	3.18	9.06	12.1	20.4	34.9	44	Weibull	16.3	13.1

N = number of samples; P05-P95 = percentiles; Pop-Estd = population-estimated; SDev = standard deviation.

Data were obtained from Tables 13-28 and 11-2 of the EFH (U.S. EPA, 1997b). Data for consumption of home-produced dairy products (in g WW/kg/d) were presented for those 20 to 39 years old and farmers (Table 13-28). Per capita intake data for dairy products (including store-bought products) were available for those younger than 1 year and those 1 to 2, 3 to 5, 6 to 11, and 12 to 19 years old (Table 11-2). Weighted averages of percentiles, means, and standard deviations were calculated for the child2 age group (combining those 1 to 2 years old and those 3 to 5 years old). Percentile data were used to fit parametric models (gamma, lognormal, and Weibull) using maximum likelihood estimation. Measures of goodness of fit were used to select Weibull as the most appropriate model in all cases.

8.4.9 Breast Milk Consumption

Age Cohort	Data Mean (mL/d)	Data SDev	Upper Percentile	Distribution	Pop-Estd Mean (mL/d)	Pop-Estd SDev (mL/d)
<1	688	ND	980	lognormal	688	1,032
1-5	ND	ND	ND	ND	ND	ND

Pop-Estd = population-estimated; SDev = standard deviation.

The data mean and upper percentile for breast milk consumption in 1- to 12-month-olds were 688 and 980 mL/d, respectively (Table 14-16, U.S. EPA, 1997b). The lognormal model was used for breast milk consumption (12-month-olds) because no percentile or related data were available. The EFH population mean for breast milk consumption was 688 mL/d, and the population standard deviation for CV=1.5 was 1,032 mL/d.

8.4.10 Beef Consumption

EFH Data–Beef Consumption (g WW/kg/d)													HWIR Distributions		
Age Cohort	N	Data Mean	Data SDev	P01	P05	P10	P25	P50	P75	P90	P95	P99	Distribution	Pop-Estd Mean	Pop-Estd SDev
1-5		ND	ND										lognormal	3.88	4.71
6-11	38	3.77	3.662		0.663	0.753	1.32	2.11	4.43	11.4	12.5		lognormal	3.88	4.71
12-19	41	1.72	1.044		0.478	0.513	0.896	1.51	2.44	3.53	3.57		gamma	1.77	1.12
farmer	182	2.63	2.644	0.27	0.394	0.585	0.896	1.64	3.25	5.39	7.51	11.3	lognormal	2.5	2.69

N = number of samples; P01-P99 = percentiles; Pop-Estd = population-estimated; SDev = standard deviation.

Home-produced beef consumption data were obtained from Table 13-36 of the EFH (U.S. EPA, 1997b). Data (in g WW/kg/d) were presented for farmers and those 6 to 11, 12 to 19, 20 to 39, and 40 to 69 years old. Percentile data were used to fit parametric models (gamma, lognormal, and Weibull) using maximum likelihood estimation. Measures of goodness of fit were used to select the most appropriate model.

Data were not available for those 1 to 2 and 3 to 5 years old. For beef consumption for those 1 to 5 years old, the lognormal model was used because, among the other age groups, it was the best-fitted model in all but one case. The population estimated mean and standard deviation for 6- to 11-year-olds were used for the analysis (normalized for body weight) and are supported by data in Table 11-3 (per-capita intake for beef, including store-bought products), which indicate that those 1 to 2, 3 to 5, and 6 to 11 years old have the highest consumption rate of beef on a g/kg/d basis.

8.4.11 Fish Consumption

EFH Data–Fish Consumption (g/d)									HWIR Distributions		
Age Cohort	N	Data Mean	Data SDev	P50	P66	P75	P90	P95	Distribution	Pop-Estd Mean	Pop-Estd SDev
1-5		ND	ND						lognormal	6.48	19.9
6-11		ND	ND						lognormal	6.48	19.9
12-19		ND	ND						lognormal	6.48	19.9
20+		ND	ND						lognormal	6.48	19.9
all ages	1,053	6.4		2	4	5.8	13	26	lognormal	6.48	19.9

N = number of samples; P50-P95 = percentiles; Pop-Estd = population-estimated; SDev = standard deviation.

Fish consumption data were obtained from Table 10-64 of the EFH (U.S. EPA, 1997b). Data (in g/d) were available only for adult freshwater anglers in Maine. Age-specific data for children were not available; children were assumed to consume the same amount of fish as adults. Percentile data were used to fit parametric models (gamma, lognormal, and Weibull) and measures of goodness of fit were used to select lognormal as the most appropriate model.

8.4.12 Soil Ingestion

EFH Data–Soil Ingestion Data and Distributions

Age Cohort	Data Mean (mg/d)	Distribution	Pop-Estd Mean (mg/d)[a]	Pop-Estd SDev (CV=0.5)	Pop-Estd SDev (CV=1)	Pop-Estd SDev (CV=1.5)[a]
1-5	100	lognormal	100	50	100	150
6-11	ND	lognormal	50	25	50	75
12-19	ND	lognormal	50	25	50	75
adult	50	lognormal	50	25	50	75

[a]HWIR distributions.
Pop-Estd = population-estimated; SDev = standard deviation.

Mean soil ingestion rates were cited as 100 mg/d for children (400 mg/d = upper percentile), 200 mg/d for children (conservative estimate), 50 mg/d for adults, and 10 g/d for pica children (Table 4-23, U.S. EPA, 1997a). No percentile data were recommended for use in the EFH. The lognormal model was used for soil consumption for all age groups. Parameter estimates were obtained by assuming CV=0.5, 1, and 1.5.

Population standard deviations based on a CV of 1.5 were used for the HWIR analysis. Adult data were used for the child3 and child4 variables.

8.4.13 Inhalation Rate

Inhalation Rate

Age Cohort	Distribution	Population-Estimated Mean (m^3/d)	Population-Estimated SDev (m^3/d)
<1	lognormal	4.5	3.15
1-5	lognormal	7.55	3.78
6-11	lognormal	11.75	3.53
12-19	lognormal	14.0	4.2
adult	lognormal	13.3	3.99

SDev = standard deviation.

No percentile data were available for the inhalation rate, and the default lognormal model was assumed. In an analysis of inhalation data, Myers et al. (RTI, 1998) found that for those younger than 3 years, CV was close to 70 percent; for other age groups, it was close to 30 percent. The lognormal distribution was fitted by using CV=70 percent for the child1 age group; CV=50 percent [(30+70)/2] for the child2 age group; and CV=30 percent for the child3, child4, and adult age groups.

8.4.14 Shower Parameters

EFH Data–Shower Parameters (minutes)													HWIR Distributions		
Parameter	**Age Cohort**	**N**	**P02**	**P05**	**P10**	**P25**	**P50**	**P75**	**P90**	**P95**	**P98**	**P99**	**Distribution**	**Pop-Estd Mean**	**Pop-Estd SDev**
showerT	all ages 1	3,547	4		5	10	15	20	30	35	50	60	gamma	16.7	9.91
cumTroom	all ages	6,661		5		15	25	40	60	90	137	255	lognormal	34	34.6

cumTroom data mean = 35.02 and data SDev = 48.8.
N = number of samples; P02-P99 = percentiles; Pop-Estd = population-estimated; SDev = standard deviation.

Percentile data for time spent taking a shower (showerT) and cumulative time spent in the bathroom (cumTroom) were provided in the EFH (U.S. EPA, 1997c). Percentile data were used to fit parametric models (gamma, lognormal, and Weibull) using maximum likelihood estimation. Measures of goodness of fit were used to select the most appropriate model for each age variable.

8.4.15 Fixed Parameters

Fixed parameters are shown in Table 8-4 along with the value selected for the HWIR analysis and data source. These constants include variables for which limited or no percentile data were provided in the EFH: exposure frequency, showering frequency, and fraction contaminated for the various media and foodstuffs considered in the HWIR analysis. Most of these values were extracted directly from the EFH. The fraction contaminated for various foodstuffs was assumed to be equivalent to the fraction of food intake that is home-produced from Table 13-71 (U.S. EPA, 1997b). The fraction of consumed trophic level 3 (T3) and trophic level 4 (T4) fish was determined from data in Table 10-66 of the EFH (U.S. EPA, 1997b), which contains the only fish consumption data reported in the handbook with an adequate species breakdown to make this distinction. Based on EPA policy, exposure duration for residents and farmers was assumed to be equivalent to the average population mobility (Table 15-176, U.S. EPA, 1997c) and an average projected U.S. life span (76.5 years) was used as human lifetime in the carcinogenic risk calculations (Table 8-1, U.S. EPA, 1997a).

The fraction contaminated for soil and drinking water was assumed to be 1 (i.e., all soil and drinking water available for consumption at a site is potentially contaminated), with actual concentrations depending on fate and transport model results. Thus, households for which the drinking water pathway was analyzed were assumed to get 100 percent of their drinking water from groundwater. Exposure frequency is set to 350 days per year in accordance with EPA policy, assuming that residents take an average of 2 weeks' vacation time away from their homes each year.

Table 8-4. Summary of HWIR Human Exposure Factor Data Used in Modeling: Constants

Model Code	Description	Units	Average	Source
BF	Event frequency (shower)	event/d	1	EFH, Table 15-176
fmbm	Fraction of fat in maternal breast milk	fraction	0.04	EFH, Table 14-16
Ffr_f	Fraction homegrown: exposed fruit (farmer)	fraction	0.328	EFH, Table 13-71
Ffr_g	Fraction homegrown: exposed fruit (gardener)	fraction	0.116	EFH, Table 13-71
Fl_f	Fraction homegrown: exposed vegetables (farmer)	fraction	0.42	EFH, Table 13-71
Fl_g	Fraction homegrown: exposed vegetables (gardener)	fraction	0.233	EFH, Table 13-71
Fpfr_f	Fraction homegrown: protected fruit (farmer)	fraction	0.03	EFH, Table 13-71
Fpfr_g	Fraction homegrown: protected fruit (gardener)	fraction	0.094	EFH, Table 13-71
Fpl_f	Fraction homegrown: protected vegetables (farmer)	fraction	0.394	EFH, Table 13-71
Fpl_g	Fraction homegrown: protected vegetables (gardener)	fraction	0.178	EFH, Table 13-71
Fr_f	Fraction homegrown: root vegetables (farmer)	fraction	0.173	EFH, Table 13-71
Fr_g	Fraction homegrown: root vegetables (gardener)	fraction	0.106	EFH, Table 13-71
Fb_f	Fraction contaminated (home-raised): beef (farmer)	fraction	0.485	EFH, Table 13-71
Fm_f	Fraction contaminated (home-produced): milk (farmer)	fraction	0.254	EFH, Table 13-71
Ff_s	Fraction contaminated fish (recreational fisher)	fraction	0.325	EFH, Table 13-71
FT3fish	Fraction of T3 fish consumed	fraction	0.36	EFH, Table 10-66
FT4fish	Fraction of T4 fish consumed	fraction	0.64	EFH, Table 10-66
Fs	Fraction contaminated: soil	fraction	1	EPA policy
Fw	Fraction contaminated: drinking water	fraction	1	EPA policy
EFr	Exposure frequency (adult resident)	d/yr	350	EPA policy
ExDur	Exposure duration for carcinogens: residents and farmers	yr	9	EFH, Table 15-176
Lifetime	Human lifetime (used in carcinogenic risk calculations)	yr	76.5	EFH, Table 8-1

Source: EFH (U.S. EPA, 1997a, 1997b, and 1997c)

8.5 Issues and Uncertainties

Many difficult choices must be made for a comprehensive risk assessment. In this section, some caveats and possible refinements are discussed and organized as follows: source data, models, regional distributions, fixed parameters, estimation methods, goodness-of-fit tests, parametric versus nonparametric approaches, and uncertainty issues.

8.5.1 Source Data Uncertainties

For most exposure factors addressed, data analyses involved fitting distributions of data summaries from the EFH (U.S. EPA, 1997a, 1997b, 1997c), in most cases by fitting distributions to selected percentiles. In our opinion, little information is lost by fitting to percentiles versus fitting to raw data. However, some believe that such analyses should always be based on raw data, synthesizing all credible sources. The question can be settled scientifically because there is a formal statistical definition of information. We suggest pursuing this activity as a scientific support activity for selected parameters. We also suggest that EPA include a broader range of percentiles in future editions of the EFH, especially for the lower (<0.50) percentiles (e.g., 0.01, 0.02, 0.03, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, 0.97, 0.98, and 0.99), to provide better data for determining the best fit.

Similarly, the percentiles for fitting the standardized age cohorts could have been obtained by fitting distributions to the original groups, generating simulated data from the fitted distributions, mixing the simulated data in proportion to the subgroup sizes, and then fitting the distributions again. Mixing proportions would best be determined by the demographics of the population (e.g., using U.S. Bureau of the Census data) of interest for the risk assessment, rather than the original study sample sizes. Resources prevented us from pursuing this option for HWIR, but it could be considered as a science support activity to test the uncertainty of the method used.

The data sets for time spent in shower [showerT] and cumulative time spent in the bathroom [cumTroom] clearly are affected by rounding and grouping of data. The fitting methods do not account for these sources of inaccuracy but could be developed and explored depending upon the significance of these input variables.

8.5.2 Model (Distribution) Uncertainty

Three standard two-parameter probability models (gamma, lognormal, and Weibull) were used for this analysis. These distributions are special cases of a three-parameter model (generalized gamma) that contains them and allows for a likelihood ratio test of the fit of the two-parameter models. Other models are possible (e.g., Myers et al., 1998), but we believe this simple setup offers a considerable improvement over using a lognormal model in all cases and is appropriate for this analysis. In support of this conclusion, the three-parameter generalized gamma module did not significantly improve on goodness of fit over the two-parameter models in 58 of 59 cases at the 5 percent level of significance.

In the few cases where fixed values were assumed because of lack of percentile data, nondegenerate probability distributions can be assigned. Although the assumption of a point estimate (i.e., mean or median) is typically used in risk assessments when data are not adequate to develop a distribution, this does imply a degree of certainty about the value when, in fact, the uncertainty in that value may be quite large. For variables that can have significant variability *and* impact on the analysis results (i.e., significantly affect risk estimates), specifying a minimum positive standard deviation would improve uncertainty estimation for the analysis. One possible approach is to assume a reasonable minimum CV (based on available data) and use a default distribution type such as lognormal or gamma. For example, except for body weight, every population CV of human exposure factors analyzed exceeds 50 percent (see Appendix 8B). Therefore, it might be reasonably assumed that an unknown CV is (at least) 50 percent for fixed parameters whose uncertainty could have a significant impact on the analysis results.

8.5.3 Methods of Estimation

The maximum likelihood estimate method of estimating uncertainty parameters is generally considered the best approach currently available for most situations. There may be room for improvement in certain cases, however. Appendix 8B shows that the maximum likelihood estimates for the means and standard deviations agree with the data means and standard deviations much better for the gamma and Weibull models than for the lognormal model. For example, note that even in cases where the lognormal model fits best, the gamma estimate of the mean is often closer to the data mean than is the lognormal mean (i.e., the lognormal maximum likelihood estimates of the mean and standard deviation can be biased). Truncation might reduce this problem, but if applied, truncated models should be fit to the data rather than fitting a model and then truncating the distribution.

8.5.4 Testing Goodness of Fit

Although they offer significant improvement in objectivity over visual estimation, goodness-of-fit tests are subject to some uncertainty that should to be considered in their application. One area of concern is our uncertainty about how the survey statistics in the EFH (U.S. EPA, 1997a, 1997b, 1997c) were calculated.

All of the statistics that have been used to assess goodness of fit here assume a random sample, which may or may not be a valid assumption for EFH data. Specifically, many of the EFH data sources are surveys that, in many cases, do not involve purely random samples. Rather, they use clustering and stratification, primarily for economic reasons. In such cases, the calculation of estimates and their standard errors, as well as test statistics, should use the survey weights and should take the study design (e.g., clustering and stratification) into account. The EFH mentions that the SAS system was used for calculation of statistics. If the SAS UNIVARIATE procedure was used to calculate percentiles, then the percentiles are unweighted.

If the random sample assumption is not valid, the likelihood ratio test used in this analysis may be more valid than the chi-square test. Valid chi-square statistics can be devised, but they require raw data and information on the survey design and weights. One way to avoid some of the difficult goodness-of-fit issues is to use empirical distributions (bootstrapping) when the raw data are available. This nonparametric approach, however, is less convenient for risk assessment simulations than using simple parametric probability models.

8.5.5 Treatment of Uncertainty

Regarding statistical treatment of uncertainty, the situation is less clear than for estimation, where there is a fairly clear consensus in favor of the maximum likelihood estimate. Relying either on the asymptotic normality of the maximum likelihood estimate or on likelihood-based contours to get parameter uncertainty distributions can be problematic when data do not exactly fit the model, which is unfortunately always the case (Box, 1976). A partial remedy for this problem was pointed out by Huber (1967); White (1982) offers a more recent and more readable account. According to White (1982), if the model is correct, then the maximum likelihood estimate is asymptotically normal, with the correct mean and variance/covariance matrix given by the so-called information matrix (i.e., by the expectation of the negative second derivatives of the log likelihood function with respect to the parameters). Still assuming the model is correct, the information matrix is equal to the variance of the score vector (the score vector is the gradient of the log likelihood function with respect to the parameters).

If the model is false, however, then the equality of these two matrices fails, and a sandwich estimator that combines them both is a better (robust) estimate of the variance/covariance matrix (White, 1982). The pursuit of this theme leads to variance estimates that absorb (are inflated by) model lack of fit and that, therefore, automatically take some account of model uncertainty as well as parameter uncertainty. This finding has led to a substantial body of recent practical statistical methodology under the name of robust sandwich estimators of variance. As far as this analysis has seen, this area of study has received little attention in the risk assessment community. In one sense, it is a treatment for an affliction mentioned by Hattis and Burmaster (1994): “The application of standard statistical methodology to a single data set will nearly always reveal only a trivial proportion of the overall uncertainty.” Regression estimates of uncertainty may be investigated as a means of addressing this problem.

8.6 References

Bickel, P.J., and K.A. Doksum. 1977. *Mathematical Statistics*. Holden-Bay, San Francisco, CA.

Box, G.E.P. 1976. Science and statistics. *Journal of the American Statistical Association*, 71(356):791-799. December

Burmaster, D.E, and K.M. Thompson. 1998. Fitting second-order parametric distributions to data using maximum likelihood estimation. *Human and Ecological Risk Assessment* 4(2):319-339.

Hattis, D., and D. Burmaster. 1994. Assessment of variability and uncertainty distributions for practical risk analyses. *Risk Analysis* 14(5):713-730.

Huber, P.J. 1967. The behavior of maximum likelihood estimates under nonstandard conditions. In: *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics*, University of California Press.

Jennrich, R.I., and R.H. Moore. 1975. Maximum likelihood estimation by nonlinear least squares. In: Statistical Computing Section Proceedings of American Statistical Association. pp. 57-65.

Jennrich, R.I., and M.L. Ralston. 1979. Fitting nonlinear models to data. *Ann Rev Biophys Bioeng* 8:195-238.

Myers, L., J. Lashley, and R. Whitmore. 1998. *Development of Statistical Distribution for Exposure Factors.* Prepared for U.S. EPA's Office of Research and Development. Research Triangle Institute, Research Triangle Park, NC.

RTI (Research Triangle Institute). 1998. Development of Statistical Distributions for Exposure Factors. (Final Report). U.S. Environmental Protection Agency, Office of Research and Development. Research Triangle Institute, RTP, NC. March 18.

U.S. Environmental Protection Agency (EPA). 1992. *Guidelines for Exposure Assessment.* Office of Research and Development, Office of Health and Environmental Assessment, Washington, DC.

U.S. Environmental Protection Agency (EPA). 1997a. *Exposure Factors Handbook, Volume I, General Factors.* EPA/600/P-95/002Fa. Office of Research and Development, Washington, DC. August.

U.S. Environmental Protection Agency (EPA). 1997b. *Exposure Factors Handbook, Volume II, Food Ingestion Factors.* EPA/600/P-95/002Fb. Office of Research and Development, Washington, DC. August.

U.S. Environmental Protection Agency (EPA). 1997c. *Exposure Factors Handbook, Volume III, Activity Factors.* EPA/600/P-95/002Fc. Office of Research and Development, Washington, DC. August.

White, H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50(1):1-25.

Appendix 8A. HWIR Exposure Factor Raw Data: Descriptive Statistics by Standardized Age Groups

Parameter	Age Cohort	N	Avg	SDev	Units	P01	P02	P05	P10	P15	P25	P50	P75	P85	P90	P95	P98	P99
beef	6-11	38	3.77	3.662	g WW/kg/d			0.663	0.753		1.32	2.11	4.43		11.4	12.5		
beef	12-19	41	1.72	1.044	g WW/kg/d			0.478	0.513		0.896	1.51	2.44		3.53	3.57		
beef	farmer	182	2.63	2.644	g WW/kg/d	0.27		0.394	0.585		0.896	1.64	3.25		5.39	7.51		11.3
bodywt	<1	356	9.102	1.287	kg			7.053	7.451	7.852	8.252	9.151	9.752	10.4	10.65	11.15		
bodywt	1-5	3,762	15.52	3.719	kg			12.5	13.1	13.45	14.03	15.26	16.67	17.58	18.32	19.45		
bodywt	6-11	1,725	30.84	9.561	kg			22.79	24.05	25.07	26.44	29.58	33.44	36.82	39.66	43.5		
bodywt	12-19	2,615	58.45	13.64	kg			43.84	46.52	48.31	50.94	56.77	63.57	68.09	71.98	79.52		
bodywt	20+	12,504	71.41	15.45	kg			52.86	55.98	58.21	61.69	69.26	78.49	84.92	89.75	97.64		
cumTroom	all ages	6,661	35.02	48.8	min			5			15	25	40		60	90	137	255
drinkH_2O	<1	403	302.7	258.5	mL/d						100.3	255.4	413.2		666.3	780.3		
drinkH_2O	1-5	3,200	697.1	401.5	mL/d	51.62		187.6	273.5		419.2	616.5	900.8		1236	1473		1917
drinkH_2O	6-11	2,405	787	417	mL/d	68		241	318		484	731	1016		1338	1556		1998
drinkH_2O	12-19	5,801	963.2	560.6	mL/d	65.15		241.4	353.8		574.4	868.5	1247		1694	2033		2693
drinkH_2O	20+	13,394	1384	721.6	mL/d	207.6		457.5	607.3		899.6	1275	1741		2260	2682		3737
expfruit	1-5	49	2.6	3.947	g WW/kg/d				0.373		1	1.82	2.64		5.41	6.07		
expfruit	6-11	68	2.52	3.496	g WW/kg/d			0.171	0.373		0.619	1.11	2.91		6.98	11.7		
expfruit	12-19	50	1.33	1.457	g WW/kg/d			0.123	0.258		0.404	0.609	2.27		3.41	4.78		
expfruit	farmer	112	2.32	2.646	g WW/kg/d	0.072		0.276	0.371		0.681	1.3	3.14		5	6.12		15.7
expfruit	home gard.	596	1.55	2.226	g WW/kg/d	0.042		0.158	0.258		0.449	0.878	1.73		3.41	5		12.9
expveg	1-5	105	2.453	2.675	g WW/kg/d			0.102	0.37		0.833	1.459	3.226		6.431	8.587		
expveg	6-11	134	1.39	2.037	g WW/kg/d			0.044	0.094		0.312	0.643	1.6		3.22	5.47		13.3

(continued)

Appendix 8A. (continued)

Parameter	Age Cohort	N	Avg	SDev	Units	P01	P02	P05	P10	P15	P25	P50	P75	P85	P90	P95	P98	P99
expveg	12-19	143	1.07	1.128	g WW/kg/d			0.029	0.142		0.304	0.656	1.46		2.35	3.78		5.67
expveg	farmer	207	2.17	2.316	g WW/kg/d			0.184	0.372		0.647	1.38	2.81		6.01	6.83		10.3
expveg	home gard.	1,361	1.57	2.029	g WW/kg/d	0.003		0.089	0.168		0.413	0.889	1.97		3.63	5.45		10.3
milk	<1	20	62.74	12.52	g WW/kg/d				0.61		24.68	45.78	91.12		136.7	170.9		
milk	1-5	40	23.71	3.838	g WW/kg/d			2.98	7.47		13.56	21.5	32.22		42.63	49.62		
milk	6-11	20	13.33	1.181	g WW/kg/d			1.81	3.54		6.72	11.88	18.58		25.38	28.76		
milk	12-19	20	6.293	0.657	g WW/kg/d			0.27	0.61		2.31	5.29	9.2		12.75	15.12		
milk	farmer	63	17.1	15.8	g WW/kg/d			0.736	3.18		9.06	12.1	20.4		34.9	44		
profruit	12-19	20	2.96	4.441	g WW/kg/d			0.16	0.283		0.393	1.23	2.84		7.44	11.4		
profruit	20+	106	5.338	7.174	g WW/kg/d			0.276	0.342		0.82	2.127	8.022		15.25	19.8		
profruit	all ages	173	5.74	8.221	g WW/kg/d	0.15		0.266	0.335		0.933	2.34	7.45		16	19.7		47.3
profruit	home gard.	146	5.9	8.422	g WW/kg/d	0.117		0.265	0.335		1.16	2.42	7.46		16	19.1		47.3
proveg	1-5	53	1.76	1.79	g WW/kg/d			0.265	0.408		0.829	1.397	2.066		3.053	6.812		
proveg	6-11	63	1.1	1.064	g WW/kg/d			0.208	0.318		0.387	0.791	1.31		2.14	3.12		
proveg	12-19	51	0.776	0.622	g WW/kg/d			0.161	0.239		0.354	0.583	0.824		1.85	2.2		
proveg	farmer	142	1.3	1.728	g WW/kg/d	0.087		0.166	0.209		0.337	0.599	1.4		3.55	5.4		9.23
proveg	home gard.	602	1.01	1.161	g WW/kg/d	0.103		0.153	0.192		0.336	0.642	1.21		2.32	3.05		6.49
rootveg	1-5	45	1.886	2.371	g WW/kg/d			0.081	0.167		0.291	0.686	2.653		5.722	7.502		
rootveg	6-11	67	1.32	1.752	g WW/kg/d			0.014	0.036		0.232	0.523	1.63		3.83	5.59		
rootveg	12-19	76	0.937	1.037	g WW/kg/d			0.008	0.068		0.269	0.565	1.37		2.26	3.32		
rootveg	farmer	136	1.39	1.469	g WW/kg/d	0.111		0.158	0.184		0.365	0.883	1.85		3.11	4.58		7.47

(continued)

Appendix 8A. (continued)

Parameter	Age Cohort	N	Avg	SDev	Units	P01	P02	P05	P10	P15	P25	P50	P75	P85	P90	P95	P98	P99
rootveg	home gard.	682	1.15	1.494	g WW/kg/d	0.005		0.036	0.117		0.258	0.674	1.5		2.81	3.64		7.47
showerT	all ages	3,547			min	3	4		5		10	15	20		30	35	50	60

Avg = average; N = number of samples; P01-P99 = percentiles; SDev = standard deviation.
Source: *Exposure Factors Handbook* (U.S. EPA, 1997a, 1997b, 1997c).

Appendix 8B. Population-Estimated Averages, Standard Deviations, and Coefficients of Variation

Parameter	Age Cohort	N	First	Data Mean	GAM Mean	LOG Mean	WEI Mean	Data SDev	GAM SDev	LOG SDev	WEI CV	Data CV	GAM CV	LOG CV	WEI CV
beef	6-11	38	lognormal	3.77	3.83	3.88	3.86	3.66	3.48	4.71	3.67	0.97	0.91	1.22	0.95
beef	12-19	41	gamma	1.72	1.77	1.82	1.76	1.04	1.12	1.41	1.07	0.61	0.64	0.78	0.61
beef	farmer	182	lognormal	2.63	2.47	2.5	2.49	2.64	2.02	2.69	2.09	1.01	0.82	1.07	0.84
bodywt	<1	356	gamma	9.1	9.09	9.09	9.04	1.29	1.23	1.25	1.31	0.14	0.14	0.14	0.14
bodywt	1-5	3,762	lognormal	15.5	15.5	15.5	15.4	3.72	2.05	2.05	2.35	0.24	0.13	0.13	0.15
bodywt	6-11	1,725	lognormal	30.8	30.7	30.7	30.4	9.56	5.94	5.96	6.87	0.31	0.19	0.19	0.23
bodywt	12-19	2,615	lognormal	58.5	58.1	58.2	57.7	13.6	10.2	10.2	11.6	0.23	0.17	0.18	0.2
bodywt	20+	12,504	lognormal	71.4	71.2	71.2	70.7	15.5	13.2	13.3	14.8	0.22	0.18	0.19	0.21
cumTroom	all ages	6,661	lognormal	35	33.6	34	33.8	48.8	27.3	34.6	29.2	1.39	0.81	1.02	0.86
drinkH_2O	<1	403	Weibull	303	304	333	302	259	271	404	261	0.85	0.89	1.21	0.86
drinkH_2O	1-5	3,200	gamma	697	698	719	698	401	406	510	390	0.58	0.58	0.71	0.56
drinkH_2O	6-11	2,405	gamma	787	787	808	787	417	430	530	408	0.53	0.55	0.66	0.52
drinkH_2O	12-19	5,801	gamma	963	965	1,000	964	561	574	739	546	0.58	0.6	0.74	0.57
drinkH_2O	20+	13,394	gamma	1,384	1,383	1,405	1,382	722	703	821	688	0.52	0.51	0.58	0.5
expfruit	1-5	49	gamma	2.6	2.25	2.46	2.25	3.95	1.89	2.91	1.84	1.52	0.84	1.18	0.82
expfruit	6-11	68	lognormal	2.52	2.63	2.78	2.63	3.5	2.9	5.12	3.16	1.39	1.1	1.84	1.2
expfruit	12-19	50	lognormal	1.33	1.43	1.54	1.44	1.46	1.44	2.44	1.51	1.1	1.01	1.59	1.05
expfruit	farmer	112	lognormal	2.32	2.24	2.36	2.24	2.65	2.1	3.33	2.18	1.14	0.94	1.41	0.97

(continued)

Appendix 8B. (continued)

Parameter	Age Cohort	N	First	Data Mean	GAM Mean	LOG Mean	WEI Mean	Data SDev	GAM SDev	LOG SDev	WEI CV	Data CV	GAM CV	LOG CV	WEI CV
expfruit	home gard.	596	lognormal	1.55	1.51	1.57	1.52	2.23	1.47	2.3	1.58	1.44	0.97	1.46	1.04
expveg	1-5	105	gamma	2.45	2.55	3.06	2.56	2.68	2.58	5.61	2.65	1.09	1.01	1.83	1.04
expveg	6-11	134	lognormal	1.39	1.4	1.64	1.39	2.04	1.66	3.95	1.81	1.47	1.19	2.41	1.3
expveg	12-19	143	gamma	1.07	1.08	1.32	1.08	1.13	1.13	2.69	1.15	1.05	1.05	2.03	1.07
expveg	farmer	207	lognormal	2.17	2.22	2.38	2.22	2.32	2.13	3.5	2.18	1.07	0.96	1.47	0.98
expveg	home gard.	1,361	Weibull	1.57	1.57	1.95	1.57	2.03	.68	4.27	1.76	1.29	1.07	2.19	1.12
fish	all ages	1,053	lognormal	6.4	5.24	6.48	5.45		8.3	19.9	9.79		1.58	3.07	1.8
milk	<1	20	Weibull	62.7		172	65.4	12.5		1025	78.7	0.2		5.96	1.2
milk	1-5	40	Weibull	23.7	23.9	25.8	23.6	3.84	16	23.4	14.3	0.16	0.67	0.91	0.61
milk	6-11	20	Weibull	13.3	13.4	14.5	13.3	1.18	9.51	14	8.7	0.09	0.71	0.97	0.65
milk	12-19	20	Weibull	6.29	6.28	8.17	6.23	0.66	5.9	14.9	5.49	0.1	0.94	1.83	0.88
milk	farmer	63	Weibull	17.1	16.4	19.8	16.3	15.8	13.9	28.3	13.1	0.92	0.85	1.43	0.8
profruit	12-19	20	lognormal	2.96	2.62	2.91	2.62	4.44	3.05	6.39	3.36	1.5	1.17	2.19	1.28
profruit	20+	106	lognormal	5.34	5.46	6.67	5.49	7.17	6.59	17.7	7.28	1.34	1.21	2.65	1.33
profruit	all ages	173	lognormal	5.74	5.76	6.5	5.7	8.22	6.83	15.9	7.46	1.43	1.19	2.44	1.31
profruit	home gard.	146	lognormal	5.9	5.78	6.63	5.75	8.42	6.72	15.7	7.29	1.43	1.16	2.37	1.27
proveg	1-5	53	lognormal	1.76	1.81	1.88	1.82	1.79	1.46	1.98	1.53	1.02	0.8	1.05	0.84
proveg	6-11	63	lognormal	1.1	1.04	1.07	1.04	1.06	0.8	1.04	0.82	0.97	0.77	0.97	0.78

(continued)

Appendix 8B. (continued)

Parameter	Age Cohort	N	First	Data Mean	GAM Mean	LOG Mean	WEI Mean	Data SDev	GAM SDev	LOG SDev	WEI CV	Data CV	GAM CV	LOG CV	WEI CV
proveg	12-19	51	lognormal	0.78	0.76	0.77	0.77	0.62	0.56	0.69	0.58	0.8	0.73	0.89	0.76
proveg	farmer	142	lognormal	1.3	1.32	1.27	1.32	1.73	1.35	1.85	1.46	1.33	1.02	1.46	1.11
proveg	home gard.	602	lognormal	1.01	1.01	1.01	1.01	1.16	0.88	1.19	0.93	1.15	0.88	1.17	0.92
rootveg	1-5	45	lognormal	1.89	1.95	2.31	1.95	2.37	2.37	6.05	2.63	1.26	1.22	2.62	1.35
rootveg	6-11	67	Weibull	1.32	1.35	2.3	1.38	1.75	1.78	10.6	2.07	1.33	1.32	4.62	1.5
rootveg	12-19	76	Weibull	0.94		1.7	0.99	1.04		5.97	1.19	1.11		3.51	1.2
rootveg	farmer	136	lognormal	1.39	1.39	1.45	1.39	1.47	1.31	2.06	1.36	1.06	0.95	1.42	0.98
rootveg	home gard.	682	Weibull	1.15	1.15	1.49	1.15	1.49	1.26	3.61	1.32	1.3	1.1	2.42	1.15
showerT	all ages	3,547	gamma		16.7	16.9	16.8		9.91	11.8	10.1		0.59	0.7	0.6

CV = coefficient of variation; CV = SDev/avg. GAM = gamma; LOG = lognormal; N = number of samples; SDev = standard deviation; WEI = Weibull.

Appendix 8C. Crystal Ball-Estimated Location and Scale Parameters

Parameter	Age Cohort	N	First	LOG Mean	LOG SDev	GAMMA SCALE ALPHA	GAMMA SHAPE BETA	WEI SCALE ALPHA	WEI SHAPE BETA
beef	6-11	38	lognormal	3.88	4.71	3.15	1.22	3.94	1.05
beef	12-19	41	gamma	1.82	1.41	0.71	2.47	1.97	1.7
beef	farmer	182	lognormal	2.5	2.69	1.66	1.49	2.64	1.19
bodywt	<1	356	gamma	9.09	1.25	0.17	54.22	9.59	8.2
bodywt	1-5	3,762	lognormal	15.5	2.05	0.27	57.14	16.36	7.75
bodywt	6-11	1,725	lognormal	30.7	5.96	1.15	26.63	33.1	5.08
bodywt	12-19	2,615	lognormal	58.2	10.2	1.78	32.76	62.32	5.75
bodywt	20+	12,504	lognormal	71.2	13.3	2.43	29.23	76.53	5.51
cumTroom	all ages	6,661	lognormal	34	34.6	22.23	1.51	35.65	1.16
drinkH_2O	<1	403	Weibull	333	404	242.5	1.25	318.6	1.16
drinkH_2O	1-5	3,200	gamma	719	510	236.55	2.95	785.66	1.86
drinkH_2O	6-11	2,405	gamma	808	530	235.09	3.35	887.76	2.02
drinkH_2O	12-19	5,801	gamma	1,000	739	341.82	2.82	1,084.4	1.83
drinkH_2O	20+	13,394	gamma	1,405	821	356.85	3.88	1,560.8	2.11
expfruit	1-5	49	gamma	2.46	2.91	1.58	1.43	2.4	1.23
expfruit	6-11	68	lognormal	2.78	5.12	3.2	0.82	2.39	0.84
expfruit	12-19	50	lognormal	1.54	2.44	1.45	0.99	1.41	0.95
expfruit	farmer	112	lognormal	2.36	3.33	1.96	1.14	2.27	1.03
expfruit	home gard.	596	lognormal	1.57	2.3	1.44	1.05	1.49	0.96
expveg	1-5	105	gamma	3.06	5.61	2.62	0.97	2.51	0.96
expveg	6-11	134	lognormal	1.64	3.95	1.97	0.71	1.21	0.78
expveg	12-19	143	gamma	1.32	2.69	1.19	0.91	1.05	0.94
expveg	farmer	207	lognormal	2.38	3.5	2.05	1.08	2.24	1.02
expveg	home gard.	1,361	Weibull	1.95	4.27	1.81	0.87	1.48	0.89
fish	all ages	1,053	lognormal	6.48	19.9	13.15	0.4	3.54	0.59
milk	<1	20	Weibull	172	1025			59.39	0.83
milk	1-5	40	Weibull	25.8	23.4	10.67	2.24	26.47	1.7
milk	6-11	20	Weibull	14.5	14	6.74	1.99	14.82	1.56
milk	12-19	20	Weibull	8.17	14.9	5.55	1.13	6.52	1.14

(continued)

Appendix 8C. (continued)

Parameter	Age Cohort	N	First	LOG Mean	LOG SDev	GAMMA SCALE ALPHA	GAMMA SHAPE BETA	WEI SCALE ALPHA	WEI SHAPE BETA
milk	farmer	63	Weibull	19.8	28.3	11.84	1.38	17.45	1.25
profruit	12-19	20	lognormal	2.91	6.39	3.56	0.74	2.28	0.79
profruit	20+	106	lognormal	6.67	17.7	7.95	0.69	4.68	0.76
profruit	all ages	173	lognormal	6.5	15.9	8.11	0.71	4.91	0.77
profruit	home gard.	146	lognormal	6.63	15.7	7.81	0.74	5.05	0.8
proveg	1-5	53	lognormal	1.88	1.98	1.17	1.54	1.93	1.2
proveg	6-11	63	lognormal	1.07	1.04	0.61	1.7	1.13	1.29
proveg	12-19	51	lognormal	0.77	0.69	0.41	1.86	0.84	1.33
proveg	farmer	142	lognormal	1.27	1.85	1.39	0.95	1.25	0.9
proveg	home gard.	602	lognormal	1.01	1.19	0.77	1.3	1.04	1.09
rootveg	1-5	45	lognormal	2.31	6.05	2.89	0.67	1.64	0.75
rootveg	6-11	67	Weibull	2.3	10.6	2.35	0.57	1.06	0.68
rootveg	12-19	76	Weibull	1.7	5.97			0.91	0.84
rootveg	farmer	136	lognormal	1.45	2.06	1.24	1.12	1.4	1.02
rootveg	home gard.	682	Weibull	1.49	3.61	1.39	0.83	1.07	0.87
showerT	all ages	3,547	gamma	16.9	11.8	5.89	2.83	18.78	1.71

LOG = lognormal; N = number of samples; SDev = standard deviation; WEI = Weibull.

Appendix 8D. Population-Estimated Location and Scale Parameters

Parameter	Age Cohort	N	First	GAM LOC	LOG LOC	WEI LOC	GAM SCALE	LOG SCALE	WEI SCALE
beef	12-19	41	gamma	0.4	0.36	0.68	0.58	0.69	0.59
beef	farmer	182	lognormal	0.65	0.53	0.97	0.72	0.88	0.84
bodywt	<1	356	gamma	2.2	2.2	2.26	0.14	0.14	0.12
bodywt	1-5	3,762	lognormal	2.73	2.73	2.79	0.13	0.13	0.13
bodywt	6-11	1,725	lognormal	3.41	3.41	3.5	0.19	0.19	0.2
bodywt	12-19	2,615	lognormal	4.05	4.05	4.13	0.17	0.17	0.17
bodywt	20+	12,504	lognormal	4.25	4.25	4.34	0.18	0.18	0.18
cumTroom	all ages	6,661	lognormal	3.26	3.17	3.57	0.71	0.84	0.86
drinkH_2O	<1	403	Weibull	5.42	5.35	5.76	0.77	0.95	0.86
drinkH_2O	1-5	3,200	gamma	6.4	6.37	6.67	0.54	0.64	0.54
drinkH_2O	6-11	2,405	gamma	6.54	6.52	6.79	0.51	0.6	0.5
drinkH_2O	12-19	5,801	gamma	6.72	6.69	6.99	0.55	0.66	0.55
drinkH_2O	20+	13,394	gamma	7.12	7.1	7.35	0.48	0.54	0.47
expfruit	1-5	49	gamma	0.55	0.46	0.88	0.73	0.94	0.82
expfruit	6-11	68	lognormal	0.57	0.28	0.87	0.89	1.22	1.2
expfruit	12-19	50	lognormal	0.01	-0.2	0.34	0.84	1.12	1.05
expfruit	farmer	112	lognormal	0.49	0.31	0.82	0.79	1.05	0.97
expfruit	home gard.	596	lognormal	0.08	-0.12	0.4	0.82	1.07	1.04
expveg	1-5	105	gamma	0.58	0.38	0.92	0.84	1.21	1.04
expveg	6-11	134	lognormal	-0.1	-0.47	0.19	0.94	1.39	1.28
expveg	12-19	143	gamma	-0.3	-0.54	0.04	0.86	1.28	1.07
expveg	farmer	207	lognormal	0.47	0.29	0.81	0.81	1.07	0.98
expveg	home gard.	1,361	Weibull	0.07	-0.21	0.39	0.88	1.32	1.12
fish	all ages	1,053	lognormal	1.03	0.7	1.27	1.12	1.53	1.69
milk	<1	20	Weibull	NA	3.35	4.08	NA	1.9	1.2
milk	1-5	40	Weibull	2.99	2.95	3.28	0.61	0.77	0.59
milk	6-11	20	Weibull	2.39	2.34	2.7	0.64	0.81	0.64
milk	12-19	20	Weibull	1.52	1.37	1.88	0.8	1.21	0.88

(continued)

Appendix 8D. (continued)

Parameter	Age Cohort	N	First	GAM LOC	LOG LOC	WEI LOC	GAM SCALE	LOG SCALE	WEI SCALE
milk	farmer	63	Weibull	2.53	2.43	2.86	0.74	1.06	0.8
profruit	12-19	20	lognormal	0.53	0.19	0.83	0.93	1.33	1.27
profruit	20+	106	lognormal	1.25	0.86	1.54	0.95	1.44	1.31
profruit	all ages	173	lognormal	1.31	0.9	1.59	0.94	1.39	1.29
profruit	home gard.	146	lognormal	1.33	0.95	1.62	0.92	1.37	1.26
proveg	1-5	53	lognormal	0.34	0.26	0.66	0.71	0.86	0.83
proveg	6-11	63	lognormal	-0.19	-0.27	0.12	0.68	0.81	0.77
proveg	12-19	51	lognormal	-0.49	-0.55	-0.18	0.66	0.76	0.75
proveg	farmer	142	lognormal	-0.08	-0.33	0.23	0.85	1.07	1.11
proveg	home gard.	602	lognormal	-0.28	-0.42	0.04	0.75	0.93	0.92
rootveg	1-5	45	lognormal	0.21	-0.19	0.49	0.95	1.44	1.33
rootveg	6-11	67	Weibull	-0.21	-0.72	0.06	1.01	1.76	1.46
rootveg	12-19	76	Weibull	NA	-0.76	-0.1	NA	1.61	1.19
rootveg	farmer	136	lognormal	0.01	-0.18	0.34	0.8	1.05	0.98
rootveg	home gard.	682	Weibull	-0.26	-0.56	0.07	0.89	1.39	1.15
showerT	all ages	3,547	gamma	2.66	2.63	2.93	0.55	0.63	0.59

GAM = gamma; LOC = location; LOG = lognormal; N = number of samples; NA = not available; WEI = Weibull.

Appendix 8E. Population-Estimated Standard Errors of Location Parameters

Parameter	Age Cohort	N	First	MLE ASY GAM LOC STD ERR	REGRESS GAM LOC STD ERR	MLE ASY LOG LOC STD ERR	REGRESS LOG LOC STD ERR	MLE ASY WEI LOC STD ERR	REGRESS WEI LOC STD ERR
beef	6-11	38	lognormal	0.157	0.252	0.159	0.207	0.169	0.272
beef	12-19	41	gamma	0.104	0.25	0.11	0.283	0.1	0.235
beef	farmer	182	lognormal	0.062	0.102	0.066	0.054	0.067	0.119
bodywt	<1	356	gamma	0.007	0.012	0.007	0.013	0.007	0.013
bodywt	1-5	3,762	lognormal	0.002	0.007	0.002	0.006	0.002	0.019
bodywt	6-11	1,725	lognormal	0.005	0.021	0.005	0.017	0.005	0.038
bodywt	12-19	2,615	lognormal	0.003	0.014	0.003	0.011	0.004	0.03
bodywt	20+	12,504	lognormal	0.002	0.012	0.002	0.009	0.002	0.028
cumTroom	all ages	6,661	lognormal	0.01	0.483	0.011	0.089	0.011	0.484
drinkH_2O	<1	403	Weibull	0.047	0.09	0.05	0.16	0.047	0.084
drinkH_2O	1-5	3,200	gamma	0.01	0.047	0.011	0.275	0.01	0.038
drinkH_2O	6-11	2,405	gamma	0.011	0.05	0.012	0.277	0.011	0.027
drinkH_2O	12-19	5,801	gamma	0.008	0.048	0.009	0.286	0.008	0.032
drinkH_2O	20+	13,394	gamma	0.004	0.025	0.005	0.085	0.004	0.044
expfruit	1-5	49	gamma	0.125	0.145	0.138	0.195	0.127	0.147
expfruit	6-11	68	lognormal	0.14	0.26	0.151	0.146	0.157	0.267
expfruit	12-19	50	lognormal	0.15	0.278	0.164	0.252	0.162	0.3
expfruit	farmer	112	lognormal	0.092	0.183	0.101	0.109	0.099	0.18
expfruit	home gard.	596	lognormal	0.041	0.311	0.045	0.082	0.046	0.241
expveg	1-5	105	gamma	0.103	0.139	0.121	0.213	0.109	0.146
expveg	6-11	134	lognormal	0.106	0.238	0.122	0.131	0.119	0.177
expveg	12-19	143	gamma	0.09	0.111	0.109	0.258	0.096	0.117
expveg	farmer	207	lognormal	0.069	0.127	0.076	0.137	0.074	0.136
expveg	home gard.	1,361	Weibull	0.03	0.106	0.037	1.22	0.033	0.102
fish	all ages	1,053	lognormal	0.051	0.149	0.058	0.072	0.061	0.123
milk	<1		Weibull				0.882		0.41
milk	1-5		Weibull		0.086		0.189		0.056
milk	6-11		Weibull		0.058		0.144		0.033
milk	12-19		Weibull		0.11		0.296		0.099

(continued)

Appendix 8E. (continued)

Parameter	Age Cohort	N	First	MLE ASY GAM LOC STD ERR	REGRESS GAM LOC STD ERR	MLE ASY LOG LOC STD ERR	REGRESS LOG LOC STD ERR	MLE ASY WEI LOC STD ERR	REGRESS WEI LOC STD ERR
milk	farmer	63	Weibull	0.111	0.2	0.137	0.388	0.109	0.191
profruit	12-19	20	lognormal	0.273	0.308	0.304	0.219	0.307	0.317
profruit	20+	106	lognormal	0.123	0.231	0.144	0.202	0.138	0.237
profruit	all ages	173	lognormal	0.094	0.219	0.108	0.164	0.106	0.207
profruit	home gard.	146	lognormal	0.1	0.226	0.116	0.188	0.112	0.217
proveg	1-5	53	lognormal	0.116	0.184	0.121	0.134	0.124	0.202
proveg	6-11	63	lognormal	0.1	0.173	0.105	0.138	0.105	0.196
proveg	12-19	51	lognormal	0.108	0.142	0.11	0.111	0.115	0.167
proveg	farmer	142	lognormal	0.089	0.234	0.091	0.114	0.101	0.238
proveg	home gard.	602	lognormal	0.037	0.174	0.039	0.051	0.04	0.176
rootveg	1-5	45	lognormal	0.191	0.252	0.22	0.188	0.216	0.257
rootveg	6-11	67	Weibull	0.169	0.172	0.221	0.274	0.193	0.169
rootveg	12-19	76	Weibull			0.19	0.491	0.147	0.174
rootveg	farmer	136	lognormal	0.083	0.144	0.092	0.116	0.09	0.152
rootveg	home gard.	682	Weibull	0.043	0.088	0.054	0.321	0.047	0.083
showerT	all ages	3,547	gamma	0.01	0.092	0.011	0.093	0.011	0.12

GAM = gamma; LOC = location; LOG = lognormal; MLE = maximum likelihood estimation; N = number of samples; REGRESS = regression; STD ERR = standard error; WEI = Weibull.

Appendix 8F. Population-Estimated Standard Errors of Scale Parameters and Estimated Correlations Between Location and Scale Parameters

Parameter	Age Cohort	N	First	MLE ASY GAM SCA STD ERR	REGRESS GAM SCA STD ERR	MLE ASY LOG SCA STD ERR	REGRESS LOG SCA STD ERR	MLE ASY WEI SCA STD ERR	REGRESS WEI SCA STD ERR	GAM CORR	LOG CORR	WEI CORR
beef	6-11	38	lognormal	0.239	0.384	0.132	0.171	0.144	0.232	-0.08	0	-0.25
beef	12-19	41	gamma	0.237	0.572	0.128	0.327	0.14	0.328	-0.06	0.01	-0.24
beef	farmer	182	lognormal	0.101	0.165	0.055	0.044	0.061	0.109	-0.02	0	-0.3
bodywt	<1	356	gamma	0.082	0.132	0.041	0.071	0.045	0.087	-0.01	0.01	-0.28
bodywt	1-5	3,762	lognormal	0.025	0.085	0.013	0.033	0.014	0.12	0	-0.01	-0.29
bodywt	6-11	1,725	lognormal	0.037	0.162	0.019	0.066	0.021	0.157	0.01	-0.02	-0.3
bodywt	12-19	2,615	lognormal	0.03	0.121	0.015	0.047	0.017	0.139	0.01	-0.01	-0.29
bodywt	20+	12,504	lognormal	0.014	0.101	0.007	0.036	0.008	0.125	0	-0.01	-0.3
cumTroom	all ages	6,661	lognormal	0.017	0.826	0.009	0.08	0.01	0.441	-0.01	-0.06	-0.31
drinkH_2O	<1	403	Weibull	0.083	0.161	0.046	0.148	0.048	0.086	-0.04	-0.11	-0.29
drinkH_2O	1-5	3,200	gamma	0.025	0.113	0.013	0.322	0.014	0.053	-0.01	-0.01	-0.3
drinkH_2O	6-11	2,405	gamma	0.029	0.128	0.016	0.347	0.017	0.041	-0.01	-0.01	-0.3
drinkH_2O	12-19	5,801	gamma	0.018	0.111	0.01	0.324	0.011	0.045	-0.01	-0.01	-0.3
drinkH_2O	20+	13,394	gamma	0.012	0.069	0.006	0.115	0.007	0.072	-0.01	-0.01	-0.29
expfruit	1-5	49	gamma	0.21	0.242	0.118	0.167	0.126	0.146	-0.06	-0.01	-0.25
expfruit	6-11	68	lognormal	0.164	0.305	0.094	0.091	0.104	0.176	-0.04	0	-0.28
expfruit	12-19	50	lognormal	0.194	0.361	0.111	0.17	0.122	0.227	-0.04	0.02	-0.28
expfruit	farmer	112	lognormal	0.126	0.251	0.071	0.076	0.08	0.144	-0.05	0	-0.26

(continued)

Appendix 8F. (continued)

Parameter	Age Cohort	N	First	MLE ASY GAM SCA STD ERR	REGRESS GAM SCA STD ERR	MLE ASY LOG SCA STD ERR	REGRESS LOG SCA STD ERR	MLE ASY WEI SCA STD ERR	REGRESS WEI SCA STD ERR	GAM CORR	LOG CORR	WEI CORR
expfruit	home gard.	596	lognormal	0.054	0.406	0.031	0.056	0.034	0.18	-0.04	0	-0.27
expveg	1-5	105	gamma	0.13	0.176	0.077	0.135	0.083	0.111	-0.05	0.01	-0.26
expveg	6-11	134	lognormal	0.113	0.254	0.065	0.07	0.072	0.107	-0.04	-0.02	-0.29
expveg	12-19	143	gamma	0.109	0.135	0.064	0.153	0.069	0.084	-0.02	-0.02	-0.3
expveg	farmer	207	lognormal	0.094	0.174	0.053	0.095	0.058	0.107	-0.03	0	-0.29
expveg	home gard.	1,361	Weibull	0.035	0.124	0.021	0.691	0.022	0.07	-0.03	-0.01	-0.3
fish	all ages	1,053	lognormal	0.061	0.179	0.037	0.046	0.037	0.074	-0.05	-0.44	-0.45
milk	<1		Weibull				0.413		0.307		-0.08	-0.26
milk	1-5		Weibull		0.18		0.194		0.076	-0.06	0.02	-0.25
milk	6-11		Weibull		0.113		0.138		0.041	-0.06	0.03	-0.25
milk	12-19		Weibull		0.154		0.195		0.092	-0.08	0.03	-0.24
milk	farmer	63	Weibull	0.175	0.315	0.103	0.292	0.108	0.189	-0.05	0	-0.27
profruit	12-19	20	lognormal	0.304	0.343	0.176	0.127	0.194	0.2	-0.05	-0.01	-0.28
profruit	20+	106	lognormal	0.131	0.246	0.078	0.11	0.086	0.147	-0.05	0.01	-0.27
profruit	all ages	173	lognormal	0.099	0.23	0.057	0.086	0.063	0.124	-0.04	0	-0.29
profruit	home gard.	146	lognormal	0.107	0.241	0.062	0.1	0.069	0.133	-0.04	0	-0.28
proveg	1-5	53	lognormal	0.197	0.313	0.106	0.117	0.119	0.194	-0.07	0	-0.24
proveg	6-11	63	lognormal	0.182	0.316	0.098	0.129	0.108	0.202	-0.03	-0.01	-0.29
proveg	12-19	51	lognormal	0.205	0.269	0.11	0.111	0.121	0.175	-0.07	0.02	-0.25

(continued)

Appendix 8F. (continued)

Parameter	Age Cohort	N	First	MLE ASY GAM SCA STD ERR	REGRESS GAM SCA STD ERR	MLE ASY LOG SCA STD ERR	REGRESS LOG SCA STD ERR	MLE ASY WEI SCA STD ERR	REGRESS WEI SCA STD ERR	GAM CORR	LOG CORR	WEI CORR
proveg	farmer	142	lognormal	0.112	0.294	0.063	0.078	0.069	0.164	-0.04	0.01	-0.29
proveg	home gard	602	lognormal	0.056	0.264	0.03	0.04	0.034	0.149	-0.04	0	-0.28
rootveg	1-5	45	lognormal	0.199	0.263	0.118	0.1	0.13	0.154	-0.05	0.01	-0.27
rootveg	6-11	67	Weibull	0.156	0.159	0.097	0.12	0.105	0.092	-0.05	0.01	-0.26
rootveg	12-19	76	Weibull	.	.	0.094	0.244	0.099	0.116	.	0.01	-0.26
rootveg	farmer	136	lognormal	0.116	0.2	0.064	0.08	0.071	0.12	-0.02	-0.01	-0.3
rootveg	home gard	682	Weibull	0.049	0.099	0.029	0.171	0.031	0.055	-0.03	-0.01	-0.29
showerT	all ages	3,547	gamma	0.024	0.216	0.013	0.108	0.014	0.156	-0.01	-0.03	-0.29

CORR = correlation; GAM = gamma; LOC = location; LOG = lognormal; MLE = maximum likelihood estimate; N = number of samples; REGRESS = regression; SCA = scale; STD ERR = standard error; WEI = Weibull.

Appendix 8G. Parameterization Model

Denote the lognormal location and scale parameters by ML and LSL, and let SL = exp(LSL). Then the lognormal variate is exp(X), where X has a normal distribution with mean ML and standard deviation SL.

Denote the gamma location and scale parameters by MG and LSG, let a = exp(LSG), and let b = [exp(MG)]/a. Then the gamma pdf is

$$f(y) = \frac{y^{a-1} \exp\left(\frac{-y}{b}\right)}{\left[\Gamma(a) b^{a}\right]} \quad \text{(AG-1)}$$

Denote the Weibull location and scale parameters by MW and LSE, let SW = exp(LSW), let a = exp(-MW/SW), and let b = 1/SW. Then the Weibull CDF is

$$F(y) = 1 - \exp[-a y^{b}] \quad \text{(AG-2)}$$