Risk Analysis, Vol. 12, No. 1, 1992

Monte Carlo Techniques for Quantitative Uncertainty Analysis in Public Health Risk Assessments

Kimberly M. Thompson,[1,2] David E. Burmaster,[1] and Edmund A.C. Crouch[3]

Received October 16, 1990; revised August 1, 1991

Most public health risk assessments assume and combine a series of average, conservative, and worst-case values to derive a conservative point estimate of risk. This procedure has major limitations. This paper demonstrates a new methodology for extended uncertainty analyses in public health risk assessments using Monte Carlo techniques. The extended method begins as do some conventional methods—with the preparation of a spreadsheet to estimate exposure and risk. This method, however, continues by modeling key inputs as random variables described by probability density functions (PDFs). Overall, the technique provides a quantitative way to estimate the probability distributions for exposure and health risks within the validity of the model used. As an example, this paper presents a simplified case study for children playing in soils contaminated with benzene and benzo(a)pyrene (BaP).

KEY WORDS: Risk assessment; Monte Carlo simulation; uncertainty analysis.

1. INTRODUCTION

Following guidance published by the U.S. Environmental Protection Agency (EPA), most public health risk assessments assume and combine a series of average, conservative, and worst-case values to derive a point estimate of risk that is presumed to be conservative and protective of public health.[1,2] The *Interim Final Human Health Evaluation Manual,*[3] the most recent guidance document from the EPA headquarters, states:

> . . . Each intake variable in the equation has a range of values. *For Superfund exposure assessments, intake variable values for a given pathway should be selected so that the combination of all intake variables results in an estimate of the reasonable maximum exposure for that pathway.* As defined previously, the reasonable maximum exposure (RME) is the maximum exposure that is reasonably expected to occur at a site. Under this approach, some intake variables may not be at their individual maximum values but when in combination with other variables will result in estimates of RME. . . . (p. 6–19, emphasis in the original)

Unfortunately, the Agency offers no further definition—either qualitative or quantitative—for the key concept of reasonable maximum exposure. The guidance does not address the amount of conservatism which should be used in risk assessment.

The current risk assessment procedures have three major limitations. First, by selecting a combination of moderate, conservative, and worst-case assumptions, risk assessors and risk managers have no way of knowing the degree of conservatism in an assessment. Since current risk assessments generally lack sufficient uncertainty analysis, risk managers and the public may have a difficulat time putting the point estimates into some kind of perspective. Second, by setting the bias high enough to swamp the uncertainty for each of many variables—but not necessarily all the variables—risk assessments may consider scenarios that will rarely (if ever) happen. Third, it is fundamentally meaningless to run

[1] Alceon Corporation, P.O. Box 2669, Cambridge, Massachusetts 02238.
[2] Present address: Harvard School of Public Health, 665 Huntington Avenue, Bldg. I Room G13B, Boston, Massachusetts 02115.
[3] Cambridge Environmental, 58 Charles Street, Cambridge, Massachusetts 02141.

0272-4332/92/0300-0053$06.50/1

traditional sensitivity analyses (e.g., to make calculations at ± 10 or ± 25% from each input value) to determine the uncertainties in the final point estimates because many of the input variables are at or near their maxima. Consequently, the current process offers comfort if the estimated risks fall below a *de minimis* value, but it offers no context for interpreting the results if the estimated risks exceed a *de minimis* value.

Monte Carlo simulation addresses the weaknesses of the current risk assessment methods identified above.[4] In extending the regular methods for public health risk assessments, Monte Carlo techniques add several steps to estimate both point values and full distributions for the exposures and risks.[5–7] These extended techniques make the analyses more informative to risk managers and members of the public by giving some perspective of the uncertainty behind the point estimates.[8]

The first step in the Monte Carlo simulation is to determine (continuous or discrete) PDFs to describe each of the variables in the uncertainty analysis. In the simulation, each of many input variables can become a random variable (*rv*) with known or estimated PDF [or equivalently, a cumulative distribution function (CDF)]. Within this framework, a variable takes on a range of values with a known probability. Some distributions, for instance, are based on normal human variability and they come into play in the uncertainty analysis because we are uncertain who the person is that will actually be following the scenario. Similarly, the soil concentrations at the site are variable and we are uncertain which part or parts of the site the person will visit. Other variables, like the Cancer Potency Factors (which are used to derive the cancer risk from the bioavailable human doses), are uncertain because of the way they are derived. Here, it is important to determine whether any correlations exist among the input variables, and, if they do, to include these correlations in the simulation. Once the exposure models and the variables and constants for the models are defined, the next step is to use a suitable software to make a large number of realizations of the set of random variables in each model. For each realization, the computer draws one random value from the appropriate distribution for each of the random variables in the model, and computes a single result. This computation is repeated a large number of times to produce complete distributions of modeled variables. Finally, the distributions can be plotted and various statistical summaries of the results can be produced to help interpret the data.

The PDFs for the final estimates are often highly non-Gaussian in shape for two reasons. First, some or all of the input variables may not have normal or even symmetric distributions. Second, the input variables enter the formulae by multiplication and division (and subsequent summation), so that even if all inputs have Gaussian distributions, the results will not.

To illustrate the application of Monte Carlo simulation to human health risk assessment, we consider a simplified case for a hypothetical site. We estimate the PDFs, CDFs, and summary statistics for the Incremental Lifetime Cancer Risk (ILCR) for one scenario involving exposures to two chemicals (benzene and BaP) found in soil for two exposure pathways: (i) inadvertent ingestion of soil and (ii) dermal contact with soil. We choose PDFs for the key input variables. By assumption, each of these distributions is statistically independent of the others. (This assumption of independence limits the analysis as discussed below.) Each of these assumptions is reasonable (or not unreasonable) in view of the current knowledge and beliefs. The resulting simplified risk assessment illustrates the strengths and weaknesses of the Monte Carlo method.

While the Monte Carlo approach has many strengths, and while it provides quantitative estimates of the distributions of the exposures and risks to people in certain situations, the results from this simplified analysis (and any simulation) are limited by many explicit and implicit assumptions. We present this simplified analysis to explore and demonstrate the approach as an extension to methods currently recommended by the EPA, not to claim that it represents an exhaustive treatment of the technique. We seek to illustrate the Monte Carlo method as applied to a simplified public health risk assessment, and we seek to extend probabilistic analyses and interpretations in such assessments.

2. HYPOTHETICAL SITE AND EXPOSED POPULATION

For this paper, we create a hypothetical site and an exposed population for analysis. Acme, a private company, owns the 500 × 600 ft site which is located at the edge of Central City. Beginning its operations in the early 1850s, Acme used and maintained 27 coke ovens and two gas holders, and produced blue gas at the site until 1945, when the buildings and equipment were demolished. From 1952–1988, Baker Company leased the southern third of the property from Acme for use as a fuel storage and tank truck depot.

Central City created a 20-acre City Park to the north of the site in 1933. In 1989, Central City asked Acme to donate or sell the whole property to them to enlarge the City Park. At first, thinking that they might develop

the site, Acme cleared the site and removed the visually stained surface soils. However, in further talks with the city last year, Acme agreed in principle to sell the property for inclusion in the park. Depending on the outcome of a site risk assessment for the surface soils on the site, Acme retains the right to limit the use of the site to activities with little or no soil contact (e.g., a parking lot with concession stands, or a swimming pool with large concrete pavilions).

Since our purpose is to illustrate the use of Monte Carlo simulation, we consider only one of the many scenarios which could be considered for this site. The scenario considers children who would play in the park extension contemplated for the old Acme/Baker property. We assume that the children (from ages 8–18 years) will spend 3 hr per day playing at the park on the site and that they visit the park 1 day per week, 20 weeks per year for 10 years. We make the conservative and simplifying assumption that the children contact the soil enough with their hands and lower arms to have a rate of soil deposition on their skin of ~1 mg/cm^2 per day, and to ingest ~50 mg of soil from the site per day. Given the uncertainties inherent in an exposure assessment, this scenario is constructed in accordance with current EPA guidelines and using conservative (or health-protective) assumptions, in the spirit of analyzing the RME case, not the absolutely worst case.

3. EXPOSURE MODELS

To estimate health effects for compounds with carcinogenic potential, we first estimate the average daily dose that a person receives in units of milligram of bioavailable chemical per kilogram of body weight per day (mg/(kg·d)), averaged over a 70-year life [abbreviated as the ADD(life)]. The scenario requires two exposure models: (i) incidental ingestion of soil and (ii) dermal contact with soil.

Table I shows the 27 variables and constants in the two exposure models and the two Cancer Potency Factors (CPFs). The first two columns of the table show the name, symbol, and units of the variable or constant. The third column indicates whether the parameter applies to the dermal contact model, the soil ingestion model, or both. The fourth column gives the point estimates for the inputs, and the fifth column shows the parameterized distribution we used for those inputs we chose to vary. The sixth column specifies the sources of each of the point estimates and distributions, and the seventh column gives the location of the point estimate in the distribution. All of the point values are reasonable in the sense that the EPA has or could readily endorse the values for a particular site. Table II shows the exposure models [used to estimate the ADD(life) values] and the risk equations.

3.1. Ingestion of Soil

In this simplified case, we consider exposures from the incidental and inadvertent ingestion of contaminated soil (i.e., we include only children who do not exhibit pica). Equation (1) in Table II shows the exposure model used to estimate the ADD(life) for inadvertent ingestion of contaminated soil.

3.2. Dermal Contact with Soil

Risk assessments often evaluate exposures from dermal contact with contaminated soils. In 1990, McKone published a new model which estimates the uptake of chemicals from a soil matrix deposited onto the skin surface.[9] In this model, the stratum corneum is the main barrier to uptake, and the amount of chemical which passes through the stratum corneum represents the bioavailable dose. The model depends on scenario specific inputs, soil properties, skin properties, and chemical properties of the soil contaminants. Although both continuous and one-time deposition versions of the model are available, we use the one-time or unit-deposition model in this simplified analysis.

The unit-deposition model derives a Personal Exposure Factor (PEF) which, when multiplied by the concentration of the chemical in the soil, estimates the average daily dose on a day of exposure. Equation (2) in Table II shows the exposure model used to estimate the ADD(life) for dermal contact with contaminated soil. This PEF is averaged over a day of exposure and is a function of 17 variables as shown in Eqs. (3)–(5) in Table II. (Note that Eqs. (3)–(5) are only given to show how the different variables are used in the model. For details about the model, see Ref. 9.)

Since this model requires 17 inputs (and creating or finding 17 different parameterized distributions is an arduous task), we performed a standard sensitivity analysis to identify the most sensitive inputs. By varying each input variable ± 10% from its nominal value while holding all the other inputs constant, we found those variables which have the greatest effect on the output when changed. If distributions for all 17 of the input variables had been available, then we would have performed a

Table I. Variables and Constants Used in Exposure Models

Name, symbol	Units	Model	Point estimate	Distribution[a]	Source	Point estimate location	Group
Scenario specific data:							
Average body weight, BW	kg	Both	47	Normal (47,8.3)	10 (11)	Mean	I
Time soil stays on skin, T	hr	Dermal	8	Normal (6,1)	10	95th percentile	III
Average body surface area, SA	m^2	Dermal	1.4	Normal (1.4,0.17)	10 (11)	Mean	I
Fraction of skin area exposed, BF	—	Dermal	0.2	Lognormal (−2.15,0.5)	10 (11)	85th percentile	III
Skin soil loading, SL	mg/cm^2	Dermal	1	Uniform (0.75,1.25)	10	Mean	I
Soil ingestion rate, SIngR	mg/d	Ingestion	50	Lognormal (3.44,0.80)	10 (12,13)	72nd percentile	III
Exposure days per week, DpW	d/wk	Both	1		10		
Exposure weeks per year, WpY	wk/yr	Both	20		10		
Exposure years per life, YpL	yr/life	Both	10		10		
Days in year, DinY	d/yr	Both	364				
Years in lifetime, YinL	yr/life	Both	70				
Soil properties:							
Soil bulk density, ρ_b	kg/m^3	Dermal	1600	Normal (1600,80)	10	Mean	I
Soil porosity, ϕ	m^3/m^3	Dermal	0.5		9		
Soil water content, Θ	m^3/m^3	Dermal	0.3		9		
Organic carbon fraction, f_{oc}	—	Dermal	0.02		9		
Human skin properties:							
Skin thickness, δ_{skin}	m	Dermal	1.5E-05		9		
Skin fat content, f_{fat}	kg/kg	Dermal	0.1		9		
Skin water content, γ	m^3/m^3	Dermal	0.3	Normal (0.30,0.05)	10	Mean	I
Boundary layer size, δ_a	m	Dermal	0.0045		9		
Chemical properties:							
K_{ow}, benzene	—	Dermal	135		9		
K_h, benzene	—	Dermal	0.224		9		
K_{ow}, BaP	—	Dermal	1.55E+06		14		
K_h, BaP	—	Dermal	2.04E−05		14		
D_{air}	m^2/s	Dermal	5E−06		9		
D_{water}	m^2/s	Dermal	5E−10		9		
Soil concentrations:							
$Cs_{benzene}$	mg/kg	Both	3.39	Lognormal (0.84,0.77)	10	95th % C.I. of mean	II
Cs_{BaP}	mg/kg	Both	29.49	Lognormal (2.81,0.68)	10	95th % C.I. of mean	II
Relative bioavailabilities:							
RBA, benzene	—	Ingestion	1		2		
RBA, BaP	—	Ingestion	0.3		2		
Cancer potency factors:							
$CPF_{benzene}$	(kg·d)/mg	Both	2.9E−02	Lognormal (−4.33,0.67)	10 (18-20)	88th percentile	IV
CPF_{BaP}	(kg·d)/mg	Both	11.5	Lognormal (−0.79,2.39)	10 (18-20)	91st percentile	IV

[a]For a normal, the mean and standard deviation are used to describe the distribution. For a lognormal, the mean and standard deviation of the underlying normal are used to describe the distribution. For a uniform, the low and high are used to describe the distribution.

Monte Carlo simulation to determine which variables should be included as stochastic. For purposes of this example, we believe the standard sensitivity analysis is sufficient.

After defining our exposure models, we needed to (i) identify point estimates for all of the model inputs, (ii) find or formulate distributions for the inputs we want to vary, and (iii) put all of the information into an appropriate simulation program. For use in both exposure models, we formulated distributions for the concentrations (mass fractions) of benzene and BaP in the site soils and the CPFs. Considering the results of the sensitivity analysis, we formulated distributions for seven of the 17 input variables of McKone's model: body weight, the time soil stays on skin, average body surface area, fraction of skin area exposed, skin soil loading, bulk density of soil, and skin water content. In addition, for the soil ingestion model, we formulated distributions for soil ingestion rates and body weight of the children. This gives a total of 12 parameterized distributions.

Table II. Exposure Model and Risk Equations[a]

Soil ingestion model used to find the ADD(life)[b]:

$$\text{ADD (life)} = \frac{Cs \cdot \text{SlngR} \cdot \text{RBA} \cdot \text{DpW} \cdot \text{WpY} \cdot \text{YpL} \cdot 10^{-6}\ \text{kg/mg}}{\text{BW} \cdot \text{DinY} \cdot \text{YinL}} \quad (1)$$

Dermal contact with soil model used to find the ADD(life)[b]:

$$\text{ADD(life)} = \frac{Cs \cdot \text{PEF} \cdot \text{DpW} \cdot \text{WpY} \cdot \text{YpL}}{\text{DinY} \cdot \text{YinL}} \quad (2)$$

where:

$$\text{PEF} = \frac{\text{SL} \cdot \text{BF} \cdot \text{SA} \cdot 0.01}{\text{BW}} \left(\frac{K_u}{(K_u + K_v)}\right)\left(1 - \exp\left(-\frac{3600\,(\rho_b + 1000 \cdot \Theta + \phi - \Theta)\,(K_u + K_v)\,T}{\text{SL} \cdot 0.01}\right)\right) \quad (3)$$

$$K_v = \frac{0.000005 \cdot K_h}{\delta_a\,(4.8 \times 10^{-4}\,\rho_b f_{oc} K_{ow} + \Theta + K_h(\phi - \Theta))} \quad (4)$$

$$\frac{1}{K_u} = \frac{\delta_{skin} f_{fat} K_{ow}}{D_{water}\,\gamma^{(4/3)}} + \frac{\text{SL} \cdot 0.01 \cdot \phi^2\,(4.8 \times 10^{-4}\,\rho_b f_{oc} K_{ow} + \Theta + K_h(\phi - \Theta))}{(\rho_b + 1000 \cdot \Theta + \phi - \Theta)\,((\phi - \Theta)^{3.33} D_{air} K_h + \Theta^{3.33} D_{water})} \quad (5)$$

Equation used to find the ILCR[c]:

$$\text{ILCR} = \text{ADD(life)} \cdot \text{CPF} \quad (6)$$

[a]See Table I for key to symbols.
[b]Average daily dose of a compound, averaged over life during which exposure occurs, in units of mg/(kg·d).
[c]Incremental lifetime cancer risk, the additional probability that a person will develop cancer during lifetime in which exposure occurs (dimensionless probability).

4. POINT ESTIMATES AND PARAMETERIZED DISTRIBUTIONS

In this paper, we use three well-known distributions to describe the key model inputs: the normal or Gaussian distribution, the lognormal distribution, and the uniform distribution. We denote random variable X with a normal distribution as $X \sim$ Normal (μ, σ), where μ and σ represent the arithmetic mean and standard deviation, respectively. Similarly, the lognormal distribution is denoted as $X \sim$ Lognormal (m, s), where m and s represent the arithmetic mean and standard deviation of the underlying normal distribution, respectively. (The underlying normal distribution is generated by taking the logarithms of the values in the distribution.) Finally, we use the notation $X \sim$ Uniform (x_1, x_2) to show that the random variable X is distributed uniformly between fixed minimum (x_1) and maximum (x_2) values.

4.1. Chemical Concentrations in the Soils

For this hypothetical site, we synthesize a data set consistent with the site history. We estimate the exposure point concentration for each chemical in the soils as the 95th percentile of the arithmetic mean of the soil data (i.e., 3.39 mg/kg for benzene and 29.49 mg/kg for BaP). Next, following the Monte Carlo framework, we fit lognormal distributions to the synthetic data for each chemical to estimate PDFs for the exposure point concentrations (where Cs represents the concentration of the chemical in the soils on the site in mg/kg): $Cs_{\text{benzene}} \sim$ Lognormal (0.84, 0.77) and $Cs_{\text{BaP}} \sim$ Lognormal (2.81, 0.68).

4.2. Cancer Potency Factors

Because of the assumptions made and the methodology used in their derivation, CPF values estimated from human or animal data are inherently uncertain values. Incorporating uncertainties into risk assessments requires careful consideration of where such uncertainties arise, methods of characterizing those uncertainties, and the results of such methodologies (e.g., the sizes of the uncertainties) in particular cases. There are many potential sources of uncertainty, including the experimental results, the epidemiological model and doses, the interspecies extrapolation, and the route extrapolation. Extending the ideas in earlier publications,[15–17] one author (EC) evaluated the EPA CPFs for benzene and BaP, and estimated the degree to which the EPA values are overly conservative (biased) and uncertain. Based on this information, we parameterize the CPFs for benzene and BaP, for use in quantitative uncertainty analyses, as lognormal distributions conditional on certain modeling assumptions. We assume that extrapolation between animals and humans is unbiased if performed on the basis of body weight. We divide the EPA point estimate by the amount of bias (the factor by which the EPA value overestimates the median) to obtain the median of the distribution. To be consistent with our notation, we find the natural logarithm of this value to describe the distribution. Similarly, we use the natural logarithm of the uncertainty associated with the EPA "standard" value as the standard deviation. The CPFs for benzene and BaP have these distributions (each in units of (mg/(kg·d))$^{-1}$): $\text{CPF}_{\text{benzene}} \sim$ Lognormal (-4.33, 0.67) and $\text{CPF}_{\text{BaP}} \sim$ Lognormal (-0.79, 2.39).

We choose the published EPA ingestion CPFs as the point estimates of the CPFs for benzene and BaP, 2.9E-02 and 11.5 (mg/(kg·d))$^{-1}$, respectively.[18–20] These

values occur at approximately the 88th and 91st percentiles of their respective distributions. We recognize that using the ingestion CPF for dermal contact is incorrect, but it is what current guidance suggests.[3]

5. RISK ASSESSMENT

In keeping with the methods recommended by the EPA[3,11] we use Eq. (6) in Table II to estimate the Incremental Lifetime Cancer Risk (ILCR) from low-dose exposure to carcinogens by compound and by pathway. The ILCR represents the additional probability that a person will develop cancer during his lifetime and is a dimensionless probability.

5.1. Simulation Results

We have now described all of the components required to perform the simulations using a spreadsheet. The algebra in the spreadsheet describe the governing equations for source strength, transport of the contaminants, exposures, and toxicities. We calculate the point estimate of risk in the usual fashion by combining the point estimates for the inputs. These point estimates of risk represent the stopping point for most risk assessments. For this example, we present the results of the simulations of the ILCRs from (i) ingestion of soil contaminated with benzene, which has a point estimate of 8.21E-10, and (ii) dermal contact with soil contaminated with BaP, which has a point estimate of 2.96E-5.

Next, we estimate distributions of health risks using Crystal Ball (Market Engineering Corporation, Denver, Colorado). In the last column of Table I, the random variables are grouped according to their anticipated effects on the output. Group I variables have symmetric distributions, and their point estimates fall at the average or median value. Group II variables have asymmetric (lognormal) distributions and their point estimates fall above the mean. Group III variables have symmetric distributions, but their point estimates fall between the 72nd and 95th percentiles. Group IV variables are the CPFs. Each of these different groups has a different qualitative effect on the distribution for exposure dose.

When only point estimates are used in the simulation, the PDFs of the results appear as lines because there is no variability in the outputs, and the point estimates for the outputs match those estimated in the spreadsheet. Figures 1a–f and 2a–f show the frequency and cumulative distributions for the ILCRs from soil ingestion of benzene and from dermal contact with BaP in soils, respectively, using (a) group I random variables, (b) group II random variables, (c) group III random variables, (d) the combination of groups I, II, and III random variables, (e) group IV random variables, and (f) the combination of groups I, II, III, and IV random variables. The solid line extending through all of the distributions shows the location of the point estimate. The area to the right of this line represents the portion of the distribution which exceeds the point estimate. Table III provides summary statistics for these simulations, including the percentile locations of the point estimate, mean, median, mode, and icosatiles.

Group I random variables (body weight, average body surface area, skin soil loading, soil bulk density, and skin water content) have symmetric (normal or uniform) distributions, and their point estimates fall at the center of the distributions. Consequently, we expect group I variables to cause lightly skewed spread around the point estimate for the outputs, with little or no lateral shift. As expected, group I variables acting jointly cause almost symmetric variation about the point estimates for the outputs (see Figs. 1a and 2a). The ILCR point estimates are located at approximately the 50th percentiles of the distributions and the mean and mode change slightly.

Group II random variables include the concentrations of benzene and BaP in soils. Because the concentrations are modeled as lognormal distributions with the arithmetic means close to the point estimates, we expect these distributions will cause the results to have lognormally shaped distributions. As expected, the distributions for group II random inputs cause moderately skewed spread in the output distributions, with a general shift of the measures of central tendency for the output distributions to the left of the point estimates (see Figs. 1b and 2b). The locations of the point estimates shift to the 69th percentile for the ingestion of benzene in soil case and they shift to the 80th percentile for the dermal contact with BaP case.

Group III random variables include the soil ingestion rate, the length of time that soil stays on the skin, and the fraction of the body exposed. For these random variables, the point estimates exceed the 70th percentile of the respective distributions and we expect to see shifts in the distributions. As expected, the distributions in Figs. 1c and 2c show dramatic shifts in the distributions toward values lower than the point estimates. This change is also apparent in the summary statistics. The locations of the point estimates shift to the 72nd percentile for the ingestion of benzene in soil case and they shift to the 94th percentile for the dermal contact with BaP case.

Figures 1d and 2d show the distributions using groups I, II, and III random variables in the simulation. As

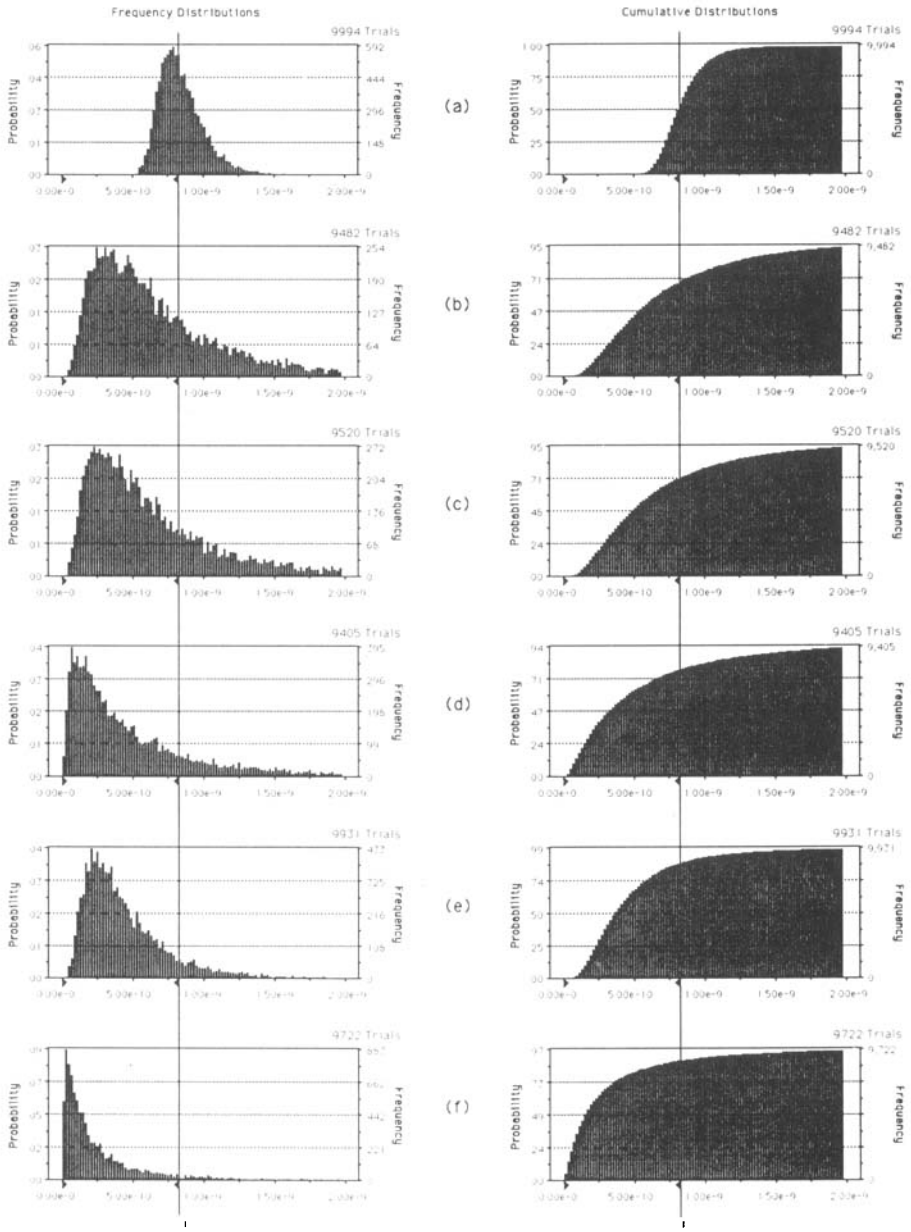

Fig. 1. Frequency and cumulative distributions for ILCR from ingestion of soils contaminated with benzene.

expected, these output distributions have long right tails, high variance, and average values much lower than the point estimates. The locations of the point estimates using all of the exposure variables shift to the 78th percentile for the ingestion of benzene in soil case and they shift to the 94th percentile for the dermal contact with BaP case.

The CPFs for benzene and BaP with the distributions given earlier are group IV random variables. Figures 1e and 2e show the distributions for the five measures of risk. As expected, we see dramatic shifts in the distributions toward values lower than the point estimates. For each of the two pathways, the point estimates fall at the 88th percentile for benzene and the 91st percentile for BaP. These simulations demonstrate the amount of conservatism built into the CPFs.

Finally, Figs. 1f and 2f show the distributions using groups I, II, III, and IV random variables in the simu-

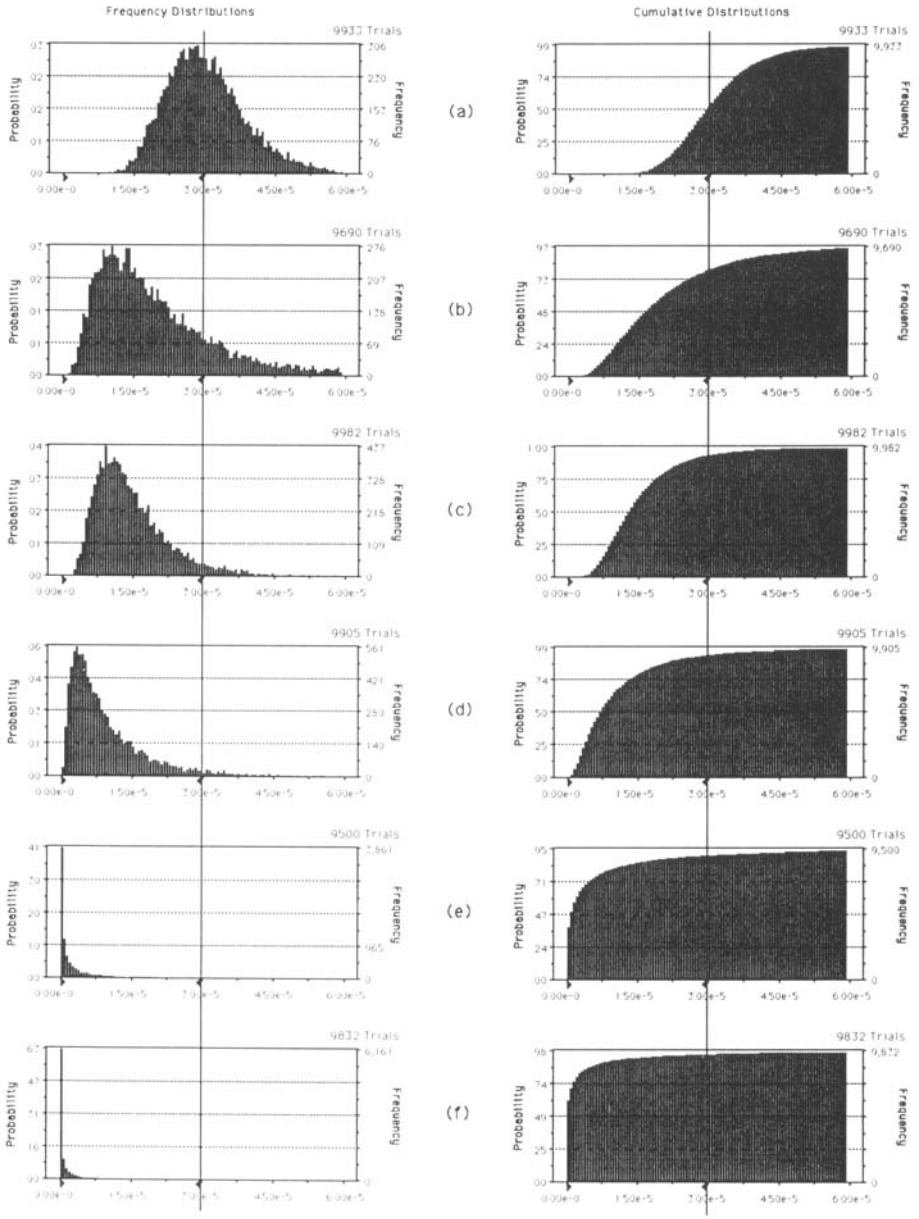

Fig. 2. Frequency and cumulative distributions for ILCR from dermal contact with soils contaminated with BaP.

lation. As expected, the output distributions in these figures have very long right tails, and most of their values are much lower than the point estimates. The point estimates of risk are at the 90th and 97th percentiles of the respective distributions. Comparing the results in the last three distributions of these figures, we see how combinations of conservative assumptions in the exposure scenarios (Figs. 1d and 2d), and in the Dose–Response Assessment (Figs. 1e and 2e) can shift the modes, medians, and even the 95th percentiles of the risk distributions far below the point estimates (Figs. 1f and 2f).

5.2. Simulation Tools, Run Times, and Convergence

A variety of companies now sell software for running Monte Carlo simulations on personal computers.[21–23] We prefer Crystal Ball, which runs in con-

Table III. Summary Statistics for Distributions Shown in Figures 1 and 2

Stochastic variable groups:	I	II	III	I,II,III	IV	I,II,III,IV
Benzene soil ingestion ILCR	(a)	(b)	(c)	(d)	(e)	(f)
Statistics:						
Point estimate location	50%	69%	72%	78%	88%	90%
Mean	8.51E−10	7.61E−10	7.17E−10	6.64E−10	4.65E−10	3.74E−10
Median (exact)	8.21E−10	5.65E−10	5.16E−10	3.58E−10	3.67E−10	1.61E−10
Mode	5.15E−10	3.54E−11	3.34E−11	6.22E−12	2.20E−11	1.25E−12
Percentile:						
0%	5.15E−10	3.54E−11	3.34E−11	6.22E−12	2.20E−11	1.25E−12
5%	6.39E−10	1.63E−10	1.42E−10	5.45E−11	1.22E−10	1.91E−11
10%	6.69E−10	2.11E−10	1.86E−10	8.15E−11	1.58E−10	3.05E−11
15%	6.95E−10	2.55E−10	2.23E−10	1.09E−10	1.84E−10	4.21E−11
20%	7.14E−10	2.98E−10	2.61E−10	1.37E−10	2.10E−10	5.46E−11
25%	7.33E−10	3.39E−10	3.00E−10	1.66E−10	2.35E−10	6.78E−11
30%	7.51E−10	3.78E−10	3.38E−10	1.95E−10	2.58E−10	8.21E−11
35%	7.68E−10	4.25E−10	3.80E−10	2.29E−10	2.85E−10	9.91E−11
40%	7.86E−10	4.70E−10	4.21E−10	2.65E−10	3.11E−10	1.18E−10
45%	8.03E−10	5.14E−10	4.68E−10	3.07E−10	3.39E−10	1.39E−10
50%	8.21E−10	5.65E−10	5.16E−10	3.58E−10	3.67E−10	1.61E−10
55%	8.39E−10	6.21E−10	5.70E−10	4.13E−10	4.01E−10	1.87E−10
60%	8.60E−10	6.84E−10	6.32E−10	4.77E−10	4.38E−10	2.24E−10
65%	8.82E−10	7.53E−10	6.98E−10	5.49E−10	4.79E−10	2.65E−10
70%	9.08E−10	8.41E−10	7.84E−10	6.46E−10	5.29E−10	3.16E−10
75%	9.34E−10	9.51E−10	8.93E−10	7.49E−10	5.82E−10	3.80E−10
80%	9.69E−10	1.08E−9	1.02E−9	8.98E−10	6.49E−10	4.76E−10
85%	1.01E−9	1.25E−9	1.20E−9	1.12E−9	7.35E−10	6.14E−10
90%	1.06E−9	1.52E−9	1.47E−9	1.47E−9	8.80E−10	8.59E−10
95%	1.16E−9	2.03E−9	1.97E−9	2.21E−9	1.13E−9	1.38E−9
100%	3.04E−9	1.10E−8	8.47E−9	1.95E−8	5.14E−9	2.29E−8
BaP dermal contact ILCR	(a)	(b)	(c)	(d)	(e)	(f)
Statistics:						
Point estimate location	51%	80%	94%	94%	91%	97%
Mean	3.06E−5	2.12E−5	1.50E−5	1.09E−5	2.10E−5	7.72E−6
Median (exact)	2.94E−5	1.68E−5	1.30E−5	7.22E−6	1.18E−6	2.87E−7
Mode (exact)	7.94E−6	1.25E−6	1.55E−6	1.90E−7	8.44E−11	1.13E−11
Percentile:						
0%	7.94E−6	1.25E−6	1.55E−6	1.90E−7	8.44E−11	1.13E−11
5%	1.80E−5	5.48E−6	5.48E−6	1.61E−6	2.37E−8	4.30E−9
10%	2.03E−5	6.95E−6	5.57E−6	2.24E−6	5.73E−8	1.06E−8
15%	2.18E−5	8.28E−6	7.60E−6	2.82E−6	1.04E−7	2.01E−8
20%	2.32E−5	9.42E−6	8.38E−6	3.36E−6	1.64E−7	3.14E−8
25%	2.43E−5	1.06E−5	9.17E−6	3.90E−6	2.43E−7	4.89E−8
30%	2.54E−5	1.17E−5	9.91E−6	4.48E−6	3.46E−7	7.19E−8
35%	2.64E−5	1.30E−5	1.07E−5	5.04E−6	4.78E−7	1.04E−7
40%	2.74E−5	1.41E−5	1.14E−5	5.70E−6	6.51E−7	1.43E−7
45%	2.84E−5	1.54E−5	1.22E−5	6.40E−6	8.82E−7	2.05E−7
50%	2.94E−5	1.68E−5	1.30E−5	7.22E−6	1.18E−6	2.87E−7
55%	3.05E−5	1.83E−5	1.39E−5	8.08E−6	1.59E−6	3.85E−7
60%	3.16E−5	2.00E−5	1.49E−5	9.07E−6	2.21E−6	5.33E−7
65%	3.28E−5	2.18E−5	1.60E−5	1.02E−5	3.02E−6	7.61E−7
70%	3.40E−5	2.40E−5	1.72E−5	1.17E−5	4.13E−6	1.08E−6
75%	3.55E−5	2.66E−5	1.85E−5	1.35E−5	6.02E−6	1.59E−6
80%	3.71E−5	2.96E−5	2.03E−5	1.56E−5	8.81E−6	2.49E−6
85%	3.93E−5	3.37E−5	2.24E−5	1.84E−5	1.39E−5	3.90E−6
90%	4.22E−5	3.97E−5	2.56E−5	2.31E−5	2.50E−5	7.50E−6
95%	4.70E−5	5.16E−5	3.11E−5	3.20E−5	5.98E−5	1.88E−5
100%	1.22E−4	2.44E−4	1.04E−4	2.50E−4	1.60E−2	1.19E−2

junction with Excel on Apple Macintosh computers. Crystal Ball (v.2) running on a 25-MHz Apple Macintosh IIci computer performed all the simulations in this paper. A simulation with 10,000 iterations takes ~16 min. We compare the results from independent simulations as a way to test the convergence and stability of the results for the highly skewed distributions shown in Figs. 1d and 2d. First, for two independent runs of 10,000 iterations each, the estimated means, standard deviations, variances, and the 90th and 95th percentiles agree within 1%. Second, for two independent runs of 10,000 and 20,000 iterations, all of the summary statistics—except the sensitive 95th percentile and maximum (which agreed to within slightly over 1%)—agreed within 1%. From this, we conclude that 10,000 iterations are sufficient to ensure convergence and stability of the output distributions. In a comprehensive Monte Carlo simulation, it is important to determine an appropriate number of iterations to support the final statements (unless the software does so automatically). It is also important to verify that the selected random number generator used by the simulation software has a sufficiently large cycle to prevent degeneracy in the simulation.

6. DISCUSSION

Advanced spreadsheets running on personal computers now provide an easy and fast way to estimate full probability distributions for human health risks in assessments conducted for sites with chemical contamination. While the methods are straight forward and can easily be extended to linked spreadsheets, and while the arrival of new software such as Crystal Ball speed the computations, more research is needed to determine and justify the specification of input distributions for exposure–related variables, and new methods are needed to quantify the distributions appropriate for Cancer Potency Factors.

As mentioned previously, this paper rests on many assumptions which simplify the analysis but which also limit the results. While it is not possible to list all the simplifications—hence limitations—it is important to discuss some of the main types and to give illustrations. First and foremost, the paper uses greatly simplified equations to estimate exposure to chemicals. While in the spirit of current federal guidance for public risk assessments, these equations are dramatic simplifications of reality. This example demonstrates the point: the equation used to estimate children's exposure to soil ignores changes in body weight and in behavior as a function of age. The equation rests on the further assumption that all children are identical in size and behavior, surely an oversimplification. Second, the paper ignores obvious correlations among variables (although Crystal Ball can handle simple correlations). As a prominent example, body weight and skin area are certainly correlated, and the joint distribution of these variables is again a function of age and sex of the child. As a less obvious example, the thickness of the skin and the water content of it are surely correlated, if not directly related. Third, even in the Monte Carlo simulations, the paper treats many variables known to be stochastic as deterministic. For example, the simulations consider that children visit the site (i) on a fixed number of days per week, (ii) for a fixed number of weeks per year, and (iii) for a fixed number of years in their lives. Surely these simplifying assumptions limit the interpretation and application of the results. While it is relatively easy to overcome the third class of oversimplifications and limitations within current knowledge and computational resources, much more research is needed to address and resolve the serious limitations imposed by the first two classes of simplifying assumptions.

ACKNOWLEDGMENTS

We are grateful to Thomas E. McKone and Charles A. Menzie for their helpful comments and suggestions. We thank the Gas Research Institute (GRI), the U.S. Environmental Protection Agency, and Alceon Corporation for partial support of this research.

REFERENCES

1. U.S. Environmental Protection Agency, *Final Report Exposure Factors Handbook* (Office of Health and Environmental Assessment, EPA/600/8–89/043, Washington, D.C., May 1989).
2. U.S. Environmental Protection Agency, Region I, *Draft Final Risk Assessment Guidance, Region I* (Boston, February 1989).
3. U.S. Environmental Protection Agency, *Risk Assessment Guidance for Superfund, Volume I, Human Health Evaluation Manual, Part A, Interim Final* (Office of Emergency and Remedial Response, EPA/540/1–89/002, Washington, D.C., December 1989).
4. R. Y. Rubenstein, *Simulation and the Monte Carlo Method* (John Wiley & Sons, New York, 1981).
5. D. E. Burmaster and K. E. von Stackelberg, "Quantitative Uncertainty Analysis in Exposure and Dose–Response Assessments in Public Health Risk Assessments Using Monte Carlo Techniques," *Proc. Natl. Conf. on Hazardous Waste and Hazardous Materials* (Washington, D.C., HMCRI, Silver Spring, November 1989), pp. 82–85.
6. D. E. Burmaster and K. E. von Stackelberg, "Monte Carlo Simulations of Uncertainties in Risk Assessments of Superfund Sites Using Crystal Ball®," *Proc. Natl. Conf. on Environmental Engineering* (American Society of Civil Engineers, July 1989).
7. D.E. Burmaster and K. E. von Stackelberg, "A New Method for

Uncertainty and Sensitivity Analysis in Public Health Risk Assessments at Hazardous Waste Sites Using Monte Carlo Techniques in a Spreadsheet," *Proc. Natl. Conf. on Hazardous Waste and Hazardous Materials* (Washington, D.C., HMCRI, Silver Spring, November 1988), pp. 550–556.
8. A. M. Finkel, *Confronting Uncertainty in Risk Management, A Guide for Decision-Makers* (Resources for the Future, Washington, D.C., 1990).
9. T. E. McKone, "Dermal Uptake of Organic Chemicals from a Soil Matrix," *Risk Analysis*, **10**, 407–419 (1990).
10. The authors of this paper, based on other sources (cited). For details about a particular distribution, contact authors.
11. GCA Corporation, *Development of Statistical Distributions or Ranges of Standard Factors Used in Exposure Assessments* (revised draft final report to U.S. EPA's Exposure Assessment Group in Washington, GCA/Technology Division, Chapel Hill, North Carolina, September 1984).
12. P. K. LaGoy, "Estimated Soil Ingestion Rates for Use in Risk Assessment," *Risk Analysis* 7, 355–359 (1987).
13. K. M. Thompson and D. E. Burmaster, "Parametric Distributions for Soil Ingestion by Children," *Risk Analysis* **11**, 339–342 (1991).
14. Mabey, W. R. *et al., Aquatic Fate Process Data for Organic Priority Pullutants, Final Report* (Office of Water Regulations and Standards, EPA/440/4–81/014, December 1982).
15. E. A. C. Crouch, R. Wilson, and L. Zeise, "The Risks of Drinking Water," *Water Resources Research* **19**, 1359–1375 (1983).
16. E. A. C. Crouch, "Uncertainties in Interspecies Extrapolations of Carcinogenicity," *Environmental Health Perspectives* **50**, 321–327 (1983).
17. E. A. C. Crouch and R. Wilson, "Regulation of Carcinogens," *Risk Analysis* **1**, 47–57 (1981).
18. U.S. Environmental Protection Agency, US EPA's Database documented in: *Integrated Risk Information System, Volume 1 and Electronic Information System* (Office of Health and Environmental Assessment, EPA/600/8–86/032a, Washington, D.C., March 1987).
19. U.S. Environmental Protection Agency, *Superfund Public Health Evaluation Manual* (Office of Emergency and Remedial Response, EPA/540/1–86/060, Washington, D.C., October 1986).
20. U.S. Environmental Protection Agency, *Health Effects Assessment of Benzo(a)Pyrene* (EPA/540/1–86/022 or PB86–134335, Cincinnati, September 1984).
21. D. E. Burmaster and E. C. Udell, "A Software Review of Crystal Ball®," *Risk Analysis* **10**, 343–345 (1990).
22. J. S. Salmento, E. S. Rubin, and A. M. Finkel, "A Review of @Risk®," *Risk Analysis* **9**, 255–257 (1989).
23. W. T. Barton, "Response from Palisade Corporation," *Risk Analysis* **9**, 259–260 (1989).