APPENDIX A

A.1 Uncertainty and Variability in HWIR99

The consideration of variability and uncertainty plays an important role in the HWIR99 development effort. Variability arises from the true heterogeneity of a parameter over space and/or over time. It is distinct from uncertainty which represents a lack of information or knowledge of a parameter or model either due to lack of data, or imprecise and/or insufficient measurements, or insufficient knowledge. In the case of HWIR99, variability and uncertainty of the measures of protection arises from the variability and uncertainty of the risk model input parameters, and the uncertainty of the risk model component predictions. The remainder of this appendix presents a short summary of the sources of uncertainty and variability in HWIR99, and a discussion of the importance of accounting separately for uncertainty and variability.

A.2 Source of Variability and Uncertainty

One of the principal sources of variability in the HWIR work is the variability of input parameters between sites. Example sources of variability include the between-site variability of the waste management characteristics such as area and volume, average spatial groundwater characteristics, climatic parameters, and number and type of receptors. Although spatial variability can also occur within sites, it is likely to be a significantly smaller contribution of the overall variability than the between-site variability.

There are a number of sources that contribute to the uncertainty in the prediction of the protective regulatory levels. These uncertainties can be generally classified as sampling and non-sampling errors. Sampling errors arise because the number of samples (n) where a parameter is measured (sampled) is less than the number of sites in the population (N). The magnitude of the sampling error is a function of the variability of the parameter, the sample size n, and the population size (N). In general, the magnitude of the sampling error will be proportional to the variability and inversely proportional to the sample size. Non-sampling errors are generally independent of the sample size and are generally more difficult to estimate. Examples of non-sampling errors include measurement errors, simulation model errors, errors due to non-probability samples, improperly defined population of interest, improper problem statements, and errors due to sampling from non-target populations (non-representativeness of samples).

The input parameters for the proposed framework are used to define the modeling scenario for a facility and can be grouped into four general classes: 1) variables that describe the characteristics of the waste management facility, including area and depth; 2) variables that describe the environmental conditions of the facility and its surroundings including hydrologic, hydrogeologic, meteorologic, and geochemical conditions at the site; 3) variables that describe the (physiologic and behavioral) exposure and response characteristics of the receptors; and 4) variables that describe the physical, chemical, and biochemical properties of the chemical constituents.

The first class of input parameters can exhibit variability, and uncertainty due to measurement errors and sampling errors. The second class of parameters can exhibit within and between-facility variability, and uncertainty due to data measurement errors, sampling errors, and potentially errors due to the collection of non-probability samples. The third class of parameters can exhibit between facility

variability, between individual receptor variability, and uncertainty due to sampling errors, measurement errors, and potentially errors due to the collection of non-probability samples, or non-representative samples. Finally, the fourth class of parameters are characterized by variability between batches, and uncertainty due to sampling and measurement error.

There are also a number of prediction model error sources that would arise in the Monte Carlo simulation of the nationwide distributions of the protection measures, including: the mechanistic model prediction of the multimedia emission source terms from the WMU; the multimedia fate and transport modules that predict the media contaminant concentrations; the exposure models that predict the receptor dose; and the effect/response models that predict the receptor impacts. Additionally, there is the potential error of improperly stating the problem.

A.3 Separating Variability And Uncertainty

Separating the effects of variability and uncertainty in estimating the nationwide probability distribution of measures of protection is important for a number of reasons. First, it permits the estimation of the uncertainty in any estimated measure of the nationwide variability of the protection measure. For example, instead of reporting the 90th percentile of the nationwide risk measure, the separation of the variability and uncertainty allows the reporting of the 95% confidence limits of the 90th percentile of the nationwide risk measure. Second, it allows the identification of sources of uncertainty that are potentially reducible so that strategies for reducing the uncertainty can be developed. Additionally, as shown in the following paragraph, it can affect the determination of whether a waste concentration meets the protection measure criteria.

The separation of uncertainty and variability can be accomplished through a two-stage Monte Carlo procedure that produces the $N_f \times N_i$ output matrix described in the previous section. How the uncertainty and variability are separated is case specific. and depends on whether the parameter is either: a) variable and certain; b) constant and uncertain; c) variable and uncertain; or d) constant and certain. To illustrate the basic elements of a two-stage Monte Carlo, and how separating variability and uncertainty can affect the regulatory limits, consider the hypothetical case where the probability distribution of the risk (R') of the nationwide receptors of concern for a given waste concentration , C_w, is lognormal so that the log of risk is normally distributed with unknown mean, μ, and known variance, σ^2:

$$R = \quad \text{Log}(R') \sim N(\mu, \sigma^2) \qquad \textbf{(A.1)}$$

Uncertainty occurs from lack of knowledge of the true mean μ as a result of sampling error. This uncertainty is represented by a normal probability distribution with known mean θ=-15 and known variance, τ^2=16:

$$\mu \sim N(\theta, \tau^2) \qquad \textbf{(A.2)}$$

The uncertainty in the mean, as described by the probability distribution function (pdf) in equation (A.2) could have been derived in a number of ways including Bayesian (DeGroot, 1970),

empirical Bayesian, or parametric bootstrap methods (Efron and Tibshirani, 1993). The variability in risk is given by σ^2=16, which for this example is the same as the uncertainty in risk as given by τ^2. The remainder of the discussion is based on the assumption that the protection measure is 90% of receptors protected for a target risk of 10^{-5}.

For this simple case, three cases are considered to illustrate the effects of incorporating and separating uncertainty from variability: 1) Uncertainty is included, and uncertainty and variability are separated; 2) uncertainty is included, but uncertainty and variability are not separated; and 3) uncertainty is not included.

In the first case, the separation of uncertainty and variability allows the description of the uncertainty for any given measure of the probability distribution describing the variability. In the HWIR case, the interest is in the uncertainty of the pth percentile of the nationwide risk, or more formally the upper Qu^{th} percentile of uncertainty of the Pv^{th} percentile of variability of the log risk R. For this case, the Monte Carlo would consist of an NxM matrix of log risk realizations. Each of the M columns would be generated by first generating a value of the uncertain mean, μ, from (A.2), and then simulating N values of R from the probability distribution given by (A.1) for the given value of the uncertain mean. For each column, an estimate of the Pv^{th} percentile of variability would be estimated. The M resulting estimates of the Pv^{th} percentiles of variability for each of the M columns would then be used to estimate the uncertainty as reflected by the Qu^{th} percentile of uncertainty of the Pv^{th} percentile of variability of the log risk R

In the second case, uncertainty is not separated from variability. As a result uncertainty cannot explicitly be described for the variability. Instead the pth percentile of the nationwide risk distribution incorporates both uncertainty and variability. For this case, the Monte Carlo simulation would involve the generation of a single N*M vector of realizations, where for each N values of R correspond to a given value of the uncertain mean, μ, from (A.2). The estimate of the pth percentile of the N*M vector of realizations would incorporate both uncertainty and variability.

Finally, in the third case, uncertainty is not included in the analysis so that the distribution of nationwide risk only includes variability. For this case, the Monte Carlo simulation involves N simulations of R using (A.1), with the mean given by θ. The estimate of the pth percentile of variability from the N simulated R values would only include variability.

Figures A.1 and A.2 show the different types of results that are obtained for the three cases, depending on how uncertainty and variability are addressed. The dashed line in Figure A.1, designated as P(u+v) corresponds to the second case. It represents the cumulative

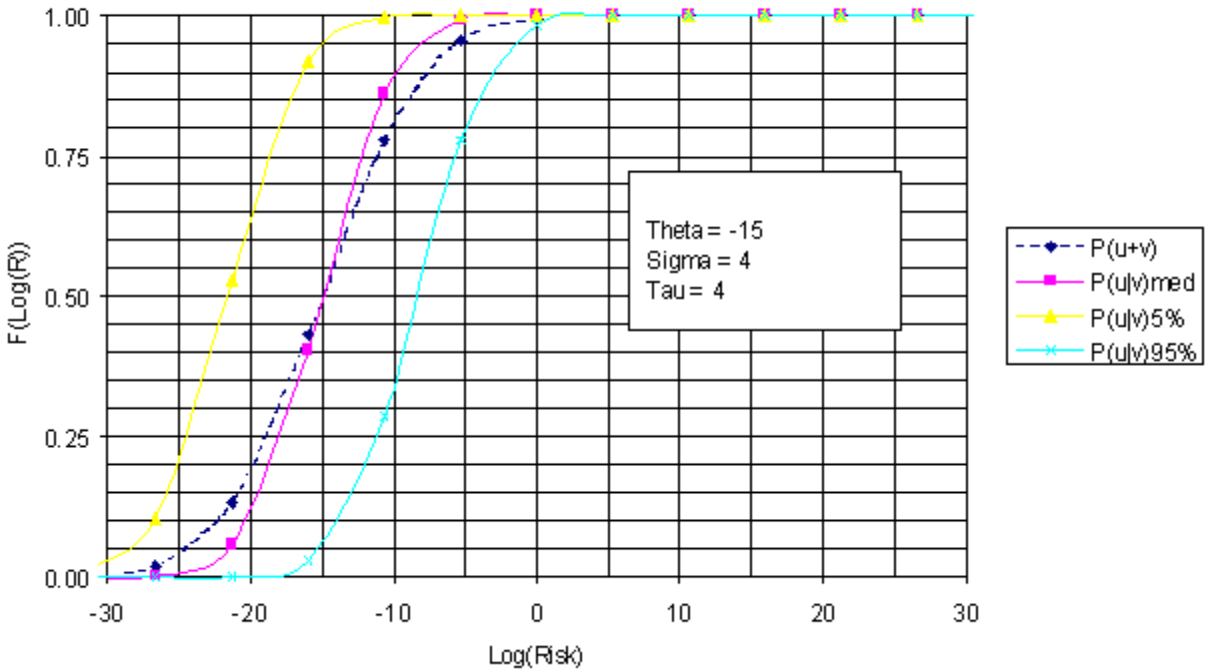

Figure A.1 Distribution of risk under uncertainty

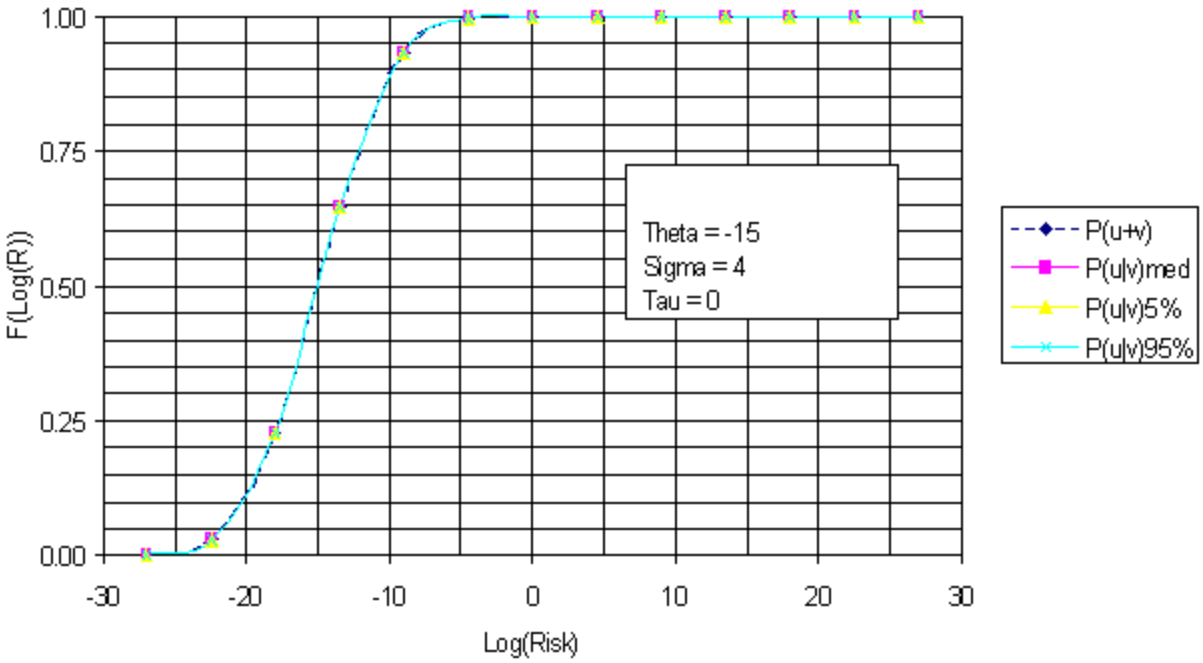

Figure A.2 Distribution of risk under no uncertainty

probability distribution function (cdf) of the log of risk for the given waste concentration based on a one-stage Monte Carlo. For a given risk value, the cdf provides an estimate of the percent of nationwide receptors whose risk is less than the given risk value. P(u+v) is obtained by analyzing the combined (NxM) output matrix of percent protections as a single data set, rather than by analyzing each iteration of the output matrix individually. As a result, the resulting cdf, P(u+v), incorporates both uncertainty and variability, but does not separate them. In particular, the resulting cdf shows that 96% of the receptors have risk less than 10^{-5}. On the basis of the one-stage Monte Carlo, the waste concentration would be considered protective of the specified protection measure.

The three curves labeled P(u|v)95%, P(u|v)5% and P(u|v)med in Figure A.1 correspond to the first case and illustrate the results of separating uncertainty and variability. Unlike the one-stage Monte Carlo, the two-stage Monte Carlo permits the estimation of the uncertainty in the protection measure by analyzing each iteration of the output matrix individually. Each iteration provides one estimate of the protection measure which can then be analyzed to estimate the uncertainty in the protection measure. The uncertainty can be depicted in a number of ways. In this example, the uncertainty is described by showing the 5% and 95% confidence limits for the cumulative distribution function of the log of risk. The curve that forms the lower envelope, and which is denoted by P(u|v)95%, indicates that there is a 95% chance that the actual percentage of protected receptors will be at least equal to the value indicated by the curve. Specifically, P(u|v)95% indicates that there is 95% chance that at least 80% of the receptors would have risk less than the target risk of 10^{-5}. Similar analysis can be used to show that there is an 89% chance that the measure of protection would be met for the given waste concentration; that the 90% receptor protection could be met with a 95% chance only for a risk of $10^{-3.3}$; and that the 96% receptor protection estimated by P(u+v) would be met with only a 77% chance. If the protection measure were modified by adding the additional constraint that the protection criteria would have to be met with at least a 95% confidence, then the waste concentration in the example would not qualify as protective.

The median curve in Figure A.1, denoted by P(u|v)med, provides an estimate of the percentage of receptors that have risk less than a specified risk if uncertainty is ignored. The same curve is shown in Figure A.2 which illustrates how the four different cdfs collapse to the median (mean) curve when the uncertainty, as represented by τ, is zero. The median (mean) curve shows that ignoring uncertainty leads to the conclusion that 99.4% of the receptors would have risk less than the target risk of 10^{-5}. Ignoring uncertainty would thus lead to accepting the waste concentration as protective.

This example illustrates the potential importance of incorporating uncertainty, and separating its effects from variability. Ignoring uncertainty and/or failing to separate uncertainty from variability prevents the characterization of the uncertainty in protection measures, and can lead to optimistic estimates of protection.

A.4 References

DeGroot, M. H., 1970. Optimal Statistical Decisions. McGraw-Hill, Inc. New York, N.Y.

Efron, Bradley and Robert J. Tibshirani, 1993. An Introduction to the Bootstrap. Chapman & Hill, New York, N.Y. 10003.