FUZZY EXPOSURE MODEL

T. RAJKUMAR and HANS W. GUESGEN
Department of Computer Science, University of Auckland
Auckland, New Zealand

ABSTRACT

This paper presents a fuzzy exposure model which deals with the uncertainties involved in analysing prolonged (chronic) chemical exposure for humans in risk assessment. The imprecise input information for the exposure model is expressed as fuzzy sets using linguistic variables such as high, low and constant. The risk assessor can extend these fuzzy sets with respect to the data availability. The result obtained from the calculations is a fuzzy number that indicates the life average daily exposure (LADE) to human beings. A case study is illustrated to present the methodology.

1. INTRODUCTION

Exposure assessment is a central component of the quantitative risk assessment procedures that are used to formulate the regulatory decisions in many countries. In its purest form, an exposure assessment initially identifies all individuals or population sub groups that have been exposed to a chemical and determines the actual doses received by the exposed individuals or populations. The calculations quantitate separate contributions to exposure by oral, dermal and inhalation routes and describe the time dependence of the exposure. Exposure assessments are based on data ranging from highly rigorous chemical analyses of an environmental contaminant to more approximate methods. The quality of data supporting an assessment and the uncertainty of the final quantitative assessment are two important and interrelated issues.

For the exposure model detailed here, hypothetical scenarios are developed which will test the model. It is recognised that the main objective of an exposure model is to provide reliable data and/or estimates for a risk assessment. Therefore it is intended that this exposure model will eventually be used to achieve this aim. Since a risk assessment requires the coupling of exposure information and toxicity or effects information, this exposure assessment process would ultimately be coordinated with toxicity/effects assessment.

2. EXPOSURE ANALYSIS

The United States Environmental Protection Agency defines exposure as "the contact with a chemical or physical agent" [1]. This implies that, while performing an assessment of human exposure to environmental contaminants, environmental concentrations are translated into quantitative estimates of the amount of contaminant that passes through the lungs, across the gut wall or through the skin surfaces of the individuals within a specified population depending on the type of exposure. This process of estimating exposure with limited data and extrapolating to large and diverse population requires many assumptions, inferences and simplifications.

The following questions should be addressed when the assumptions are made in defining the model:

- How many subgroups require modelling?
- How many age groups in each receptor population require modelling?
- What is the frequency and amount of contact between the receptor population and the contaminated media?
- How does exposure to each chemical from the various routes relate to the tissue levels of the chemical already existing in the target population (route specific [(eg) air, food,water] bio availability factors)?

For chemicals present in any medium, a dose is not easily calculated even if chemical concentration and consumption are known. Here dose refers to the concentration of the chemical affecting the human health. Absorption of the chemical may vary depending on other food constituents present, the age of the person exposed and the form of the chemical. Representative sampling is complex and analysis of very low concentrations of chemicals in environmental samples is expensive and difficult. An initial exposure assessment may be based on limited data, such as estimated ranges for input variables for an exposure prediction model. Normally the exposure assessment is site specific. It is intended that, in the absence of available data, default values can be used. These default values are expressed in three ways: averages, ranges and distributions.

0-7803-3280-6/96/$5.00 ©1996 IEEE

The basic cause of uncertainty is a lack of information available to the analyst because of inadequate, or even nonexistent, experimental and operational data on processes and parameters. Specific causes of uncertainty include the following:

- Measurement error - uncertainty arises from random and systematic error in the measurement technique
- Sampling error - uncertainty arises from representativeness of the data of the actual population being sampled
- Variability - nature's variability
- Application and quality of generic or indirect empirical data - uncertainty arises from both the applicability of the indirect data and the measurement or sampling error in the data
- Expert judgement - frequently, data gaps must be filled based on engineering or scientific assumptions, which have inherent uncertainty

Uncertainty analysis can be used to provide decision makers with a complete spectrum of information concerning the assessment and its quality. In the following sections, various methods for deriving life average daily exposure are explained in detail on the basis of availability of data and the quality of analysis.

3. CONVENTIONAL EXPOSURE METHODS

In the following two sections conventional means of determining lifetime average daily exposure on the basis of variables which are uncertain in nature are discussed.

Sensitivity Analysis

For cancer risk assessment, exposure is averaged over the body weight and lifetime.

$$LADE = \frac{\text{Total Exposure}}{\text{Body Weight} \times \text{Life Time}} \qquad (1)$$

where LADE is the lifetime average daily exposure which is expressed in mg/kgday. Total exposure is calculated as the product of contaminant concentration, contact rate and exposure duration. Contaminant concentration refers to the concentration available in the medium (air, water and soil). The contact rate refers to the rates of inhalation, dermal and oral contact.

A sensitivity analysis of exposure would estimate the range of exposures that would result as an individual model input variables were varied from their minimum to maximum and other values remain as the mid ranges. It involves the determination of the changes in model response as a result of the changes in individual parameters and tests the sensitivity of an output variable to the possible variation in the input variables of a given model. By identifying the most critical input variable (ie the one having most influence on the outcome), more resources can be directed towards identifying the underlying factors affecting the uncertainity of this input variable.

Sensitivity analysis should be performed when limited information is available about the input variables. There are two types of sensitivity analysis:

- Point sensitivity
- Range sensitivity

Point sensitivity is the sensitivity of the output to the centrality variability (mean, median, mode) of a given input variable with the other input variables held at best estimates of their centralities. Range sensitivity analysis estimates the range output values that would result as individual input variables are varied from their minimum to their maximum possible value with other input variables held at fixed values eg. mid ranges.

Probability Analysis

Sensitivity analysis can be enhanced by computing the predicted exposures using the continuous distributions as input. The monte carlo simulation is used to generate random values for input variables. A probability distribution for the input variable can be assigned based on theoretical and/or expert opinions. Unless specific information on the relationships between exposure variables is available, values for the required input parameters are assumed to be independent.

A probability distribution can be ascertained from the general shape of the distribution, minimum, maximum, mode, mean, midrange and percentiles. When distribution data is not available, expert judgement can be made on particular input variables. The following criteria is adopted for selecting the distributions:

- Uniform distribution - finite ranges of data known for the variable. The use of uniform distribution treats all values equally
- Triangular distribution - range and mode of the variable is known
- Lognormal distribution - this type of distribution is flexible in shape and provides reasonable approximations to data which are nearly symmetric as well as to some positively skewed distributions. However the mean and standard deviation are very sensitive to values of the largest observations

- Normal distributions - range of possible values symmetrically distributed around its mode

When using monte carlo simulation to produce distributions of exposures, it is important to note that standard deviation and expected value of estimated exposure of simulated values will be relatively higher than the values based on a single estimate based on the mean value of the input parameters. This is further explained in the section 5 using a case study.

4. FUZZY EXPOSURE MODEL

The fuzzy inference of the exposure model is the actual process of mapping from the input variables to an output variable using fuzzy logic. The use of fuzzy set theory will allow the user to include the unavoidable imprecision of the site data. In this paper the output variable refers to life average daily exposure and the input variables refer to the chemical concentration, duration of exposure and dose absorbed. J.F.Baldwin et.al[6] expressed advection and dispersion terms as two fuzzy sets namely high and low in the ground water model. Similar concept of linguistic sets are designed for the fuzzy exposure model.

Depending upon the data and uncertainty, the inputs are fuzzified in terms of linguistic sets (high, low etc). The output variable has to be fuzzified in a similar manner to the inputs. Rules for the model should be formulated based on the experts knowledge in the particular domain or on the basis of a literature survey. They should be realistic in terms of a particular location. The fuzzy operator is applied to obtain one number that represents the result of the antecedent for the rule. The AND is supported by minimum and OR by maximum. The implication shapes the consequent fuzzy set based on the antecedent single number. Aggregation is done through joining the output of the rules by a thread. The implication and aggregation operations are supported by minimum and maximum functions. The input for the defuzzification process is a fuzzy set and the output is the life time average daily exposure. The centroid method is adopted for finding the centre of the area under the curve which represents the LADE.

5. CASE STUDY

In this case study, life average daily ingestion exposure of an adult for a particular chemical through fish consumption is considered. The following case study is based on information obtained from the Exposure Factors Handbook [1]. The range the body weight of an adult is taken from the 5^{th} to the 50^{th} percentile. Other variables ranges are considered from the 50^{th} to 95^{th} percentile. Normally the chemical concentration in the fish would be a range of values. The diet fraction indicates the amount of chemical transferred from the fish consumed to human when the fish is being consumed. The LADE is calculated using equation (2) below. Table 1 describes the input variables and the range of values that have been chosen for the case study.

$$LADE = \frac{CR \times CC \times ED \times DF}{BW \times LT \times 365} \tag{2}$$

Table 1 Input values and distributions

Parameter	Mean	Range	Distribution
CR(consumption rate) g/day	30	30-140	Lognormal
CC (chemical concen.) mg/g	10	10	Constant
ED(exposure duration) day	3285	3285-10950	Lognormal
DF(diet fraction)	0.2	0.2-0.75	Lognormal
BW (body weight) kg	70	46.8-70	Normal
LT(lifetime) yr	75	75	Constant

Sensitivity Analysis

A sensitivity analysis has been carried out as discussed in section 2, the following results were obtained.

Table 2 Results of Sensitivity Analysis

Parameter	Estimated Exposure
CR	0.1028-0.48
CC	Fixed
ED	0.1028-0.3427
DF	0.1028-0.3855
BW	0.1028-0.1538
LT	Fixed

Probability Analysis

In this analysis monte carlo simulation was used to generate 1000 random values for the input variables as per the distribution assigned in table 1. The results of probability analysis are represented in table 3 and on figure 1 respectively.

Table 3 Results of Probability Analysis Estimated LADE (mg/kgday)

Mean	2.239
Median	1.086
Standard deviation	2.395
Range	10.36
Minimum	0.005
Maximum	10.36

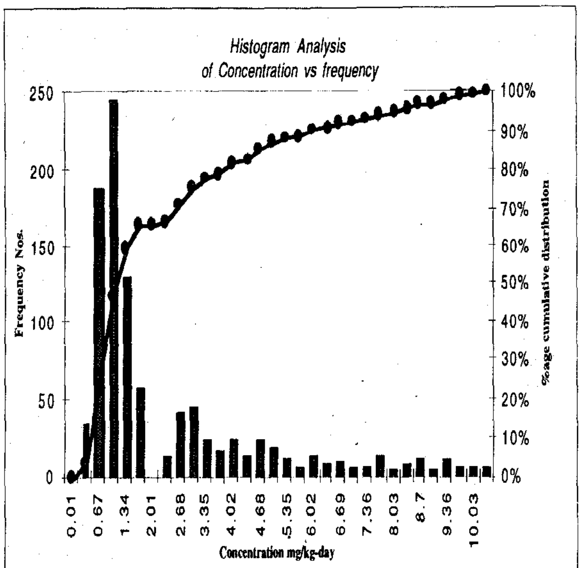

Fig 1 Histogram Analysis of simulated exposure data

Fuzzy Exposure Analysis

The risk assessor assigns the fuzzy set as per the uncertainty of the data (given in Table 1). The membership function has been assigned in a similar manner to the distributions that have been assigned to the input variables. Since the output distribution is not known clearly, a triangular membership function is assigned. In table 4 the input variables and their associated fuzzy sets are shown. Depending upon the data availability, the user may assign more fuzzy sets for input and output variables.

Table 4 Input and output data for fuzzy sets

Variable	Range	Mean	sigma	Membership function
CR	0-140	30	7.5, 55	Lognormal
CC	10	10		Constant
ED	0-14000	3285	821.5, 832.5	Lognormal
DF	0-2	0.2	0.05, 0.275	Lognormal
BW	0-110	70	11.6	Normal
LT	75	75		Constant
LADE	0-10	5		Triangular

If the above data represents a fuzzy set called 'HIGH', then 'LOW' may be defined as '1 -HIGH'. If more data is available, then 'HIGH' itself can be defined as very high, medium high and low high and vice versa. For the fuzzy exposure model, the following rules are defined to derive the final life average daily exposure.

Table 5 Rules for fuzzy inference

Rule	BW	LT	CR	CC	ED	DF	lade
1	High	Cons	Low	Cons	Low	Low	Low
2	High	Cons	High	Cons	Low	High	High
3	Low	Cons	Low	Cons	Low	Low	Low
4	Low	Cons	High	Cons	Low	Low	High
5	Low	Cons	High	Cons	Low	Low	Low
6	Low	Cons	High	Cons	High	Low	High
7	High	Cons	Low	Cons	High	High	Low

The fuzzy inference has been performed with the help of above set of rules and life average daily exposure for the above case is estimated as 4.8 mg/kgday. This is higher than the LADE obtained by using the conventional probability analysis as discussed in section 2.

6 CONCLUSION

In the sensitivity analysis, the exposure is calculated to be 0.103 mg/kg-day. In probability analysis, the mean using 1000 random input variables is calculated to be 2.24 mg/kg-day. In the histogram, most of the data generated lies between 0 to 3 mg/kg-day and frequency of events in the other ranges remain negligible. It may not be the real situation and in a probability analysis the data considered will be between the 5th and 95th percentile. In fuzzy exposure analysis, every possible realistic range is considered from 0 to maximum ranges and the fuzzy sets are defined and the defuzzification is based on the entire area of the final curve. The final fuzzy result of the model represents the life average daily exposure of a human for a particular set of data and rules for a specific site. The fuzzy exposure model is highly flexible; as more data becomes available the structure of the model can easily be modified. In fact if the user finds the results unacceptable, the model can be altered to suit the user needs. Additionally the fuzzy exposure model presented here utilises relatively inexpensive, imprecise and/or subjective data. The fuzzy exposure model analyses one specific route at one time. Continuing research is being directed towards analysing every combination of routes of exposure and effects modelling using fuzzy logic together.

ACKNOWLEDGMENT

The authors wish to thank Mr Gavin Fisher and Ms Janet Petersen of National Institute of Water and Atmospheric Research Ltd, Auckland for valuable

suggestions and comments during the development of fuzzy exposure model.

REFERENCES

[1] United States EPA, "Exposure Factors Hand Book" 1988

[2] United States EPA, "The Risk Assessment Guidelines", 1986

[3] John L schaum, John L segna, John S Young, Carol M Benes and Warren R Muir, "Computerised system for performing risk assessments", ACS symposium series 431, 1990

[4] Yong W Lee, Mohamed F Dahab and Istvan Bogardi, "Nitrate risk assessment using fuzzy set approach", Journal of Environmental Engineering, 1995

[5] Michael E Hagemeister David D Jones and Wayne E Woldt, "Hazard Ranking of landfills using fuzzy composite programming", Journal of Environmental engineering, 1996

[6] J F Baldwin, T.P. Martin & Y.Zhou "A fril Knowledge base for the management of uncertainty in performance assessment of hazardous waste repositories", Second IEEE international conference on fuzzy systems , 1993

[7] David J severn, "Exposure assessment", Environment Science Technology, 1159 - 1163, 1987