Journal of the Air & Waste Management Association

ISSN: 1096-2247 (Print) 2162-2906 (Online) Journal homepage: http://www.tandfonline.com/loi/uawm20

Health Risk Assessment of Fluctuating Concentrations Using Lognormal Models

Bernard E. Saltzman

To cite this article: Bernard E. Saltzman (1997) Health Risk Assessment of Fluctuating Concentrations Using Lognormal Models, Journal of the Air & Waste Management Association, 47:11, 1152-1160, DOI: 10.1080/10473289.1997.10464064

To link to this article: https://doi.org/10.1080/10473289.1997.10464064

TECHNICAL PAPER

ISSN 1047-3289 *J. Air & Waste Manage. Assoc.* **47**:1152-1160

Health Risk Assessment of Fluctuating Concentrations Using Lognormal Models

Bernard E. Saltzman

Department of Environmental Health, University of Cincinnati, Cincinnati, Ohio

ABSTRACT

A mathematical model is proposed for assessing health risk rates of fluctuating concentrations. Each time-averaged concentration may be regarded as a dose that, when applied to the dose-response curve, produces a risk of an adverse effect. A theoretical derivation shows that the dose-response pattern is a cumulative lognormal curve because of the diversity of the individuals in the exposed population. Similarly, the concentration pattern is a lognormal distribution because of the diversity of emission sources and dispersive processes. The health risk is produced by the overlapping of the right tail of the concentration distribution and the left tail of the dose-response curve. The evaluation of the joint probability in this region has been performed by numerical integration by computer in terms of two generalized parameters. One represents the geometric standard deviation of the concentration distribution relative to that of the dose-response curve, and the other represents the distance between the geometric mean concentration and the concentration producing an adverse response in 50% of the exposed population. These results are presented graphically and in tabular form. If the two parameters of the dose-response curve are known, the health risk of the concentration pattern may be calculated conveniently for any geometric mean and geometric standard deviation values.

IMPLICATIONS

Assessment of health risk rates is necessary to establish adequate, but not excessively low, allowable levels of pollutants. Some current levels have been set using safety factors as large as 10^5 because of lack of data and a good theoretical model. Calculated risks of exceeding the allowable level are not health risk rates. The proposed model clarifies the patterns of dose-response curves and concentration distributions that may be expected. It also clarifies exactly the adverse effects that may be expected for various geometric mean concentrations and geometric standard deviation values.

INTRODUCTION

The rising costs of environmental controls have increased concern about whether available resources are being put to their best use. Control policies must be formulated with consideration to costs, health risk rates, and benefits from avoiding these risks. Issues of assessing environmental health risks have been reviewed by the National Research Council,[1] the Office of Technology Assessment of the U.S. Congress,[2] and in a collection of case studies.[3] The present status of this assessment is limited by the lack of good data and of a good theoretical model based on scientific principles. Many control levels have been set on the basis of toxicological data, usually for animals, giving the lowest level at which adverse effects are observed (LOAEL) or the highest level at which no adverse effects are observed (NOAEL). It should be noted that these levels might be lower if more animals or people were studied. Thus, they are really estimates of levels giving a low probability of an adverse effect, rather than of zero probability. Arbitrary safety factors as high as 10^5 have been applied to these levels, probably making the recommended control levels unnecessarily stringent and expensive. Probabilities of exceeding these levels may be calculated, but they are not the same as actual risk rates of adverse health effects.

Health risk is really a rate of adverse effects during a stated period of time. To evaluate health risk rates, it is essential to have a dose-response model. Data for these are very scanty and are usually based upon laboratory exposures to a series of constant concentrations at different levels. Complications arise from the fact that environmental concentrations are not constant, but fluctuating. A linear dose-response model[4] has been used for carcinogens with a risk level of 10^{-6}. This is the most conservative of several common mathematical models. For a linear model, the average concentration may be used for risk calculations. Other models, in order of decreasing estimates of risk at extrapolated very low doses, are the multistage, Weibull, logit, multi-hit and probit models.[5-8] Mathematical models have been extensively reviewed.[9]

For non-linear models, higher concentrations give disproportionately higher effects; the mean of a fluctuating concentration will therefore not give an accurate risk value. The highest concentration observed during a stated period (or likely to be observed with a stated probability) may be applied to the dose-response curve, to give the maximum risk rate during that period. However, the health risk rate will fluctuate during the period. It should be noted that a high risk rate for a very brief time may be less dangerous than a moderate risk rate for a large fraction of the time interval during which concentrations are observed.

Time-averaged concentrations may be regarded as varying instantaneous doses. The risk at each moment is the ordinate of the dose-response curve at the effective concentration dose. If the risks are very low, they may be added for a period of time to determine the health risk rate. The purpose of this study is to develop a practical method of doing such a calculation using theoretical models based on scientific principles. The health risk is produced by the overlapping of the right tail of the concentration distribution and the left tail of the dose-response curve. The integration of the joint probability in this region has been performed by numerical integration by computer in terms of two generalized parameters. The results are presented in graphical and tabular form to permit the assessment of health risks without need of a computer.

THEORETICAL MODELS

To calculate health risk rates, three theoretical models are needed: the statistical distribution pattern of the fluctuating concentrations, the dose-response curve, and a formula for integrating the risks resulting from their interactions. The symbols used are listed in the nomenclature section. Illustrative calculations will be described. Full details of the formulas are given in the appendix.

Lognormal Model for Distributions of Concentrations

Concentrations have been well known to follow a lognormal pattern. This observed distribution can be readily justified on a theoretical basis. Each observed concentration, C_i, may be equated to the median concentration, C_m, multiplied by a series of N random factors, Mj_i, each relative to its median value:

$$C_i = C_m\, M1_i\, M2_i\, M3_i \ldots\ldots MN_i \qquad (1)$$

Thus, $M1_i$ may be a function of the number of operating pollution sources divided by the median value of this function, while $M2_i$ may be a relative function of wind velocity; $M3_i$ of wind direction; $M4_i$ of time of day; $M5_i$ of season; $M6_i$ of location; $M7_i$ of measurement errors; and so on. Each of these relative values may differ randomly for each observed concentration.

Taking logarithms of both sides of the equation, it becomes:

$$\log(C_i) = \log(C_m) + \log(M1_i) + \log(M2_i) + \log(M3_i) + \ldots\ldots \log(MN_i) \qquad (2)$$

Thus, the logarithms of the observed concentrations can be represented as a constant value plus the sums of N random variables. According to the central limit theorem of statistics, the distribution pattern of observations of a variable that is the sum of a large number of random variables approaches a normal distribution. This is the basis of most widely applied statistics. Values of $\log(C_i)$ therefore approach a normal distribution pattern. Consequently, by definition, the values of C_i will approach a lognormal distribution pattern. In general, whenever observed variables are affected by many additive random factors, a large set of observations will follow a normal distribution, and when they are affected by many multiplicative random factors, it will follow a lognormal distribution pattern.

The lognormal distribution has no negative values. This is a more natural pattern than that of the normal distribution, which theoretically has negative values. Lognormal distributions are commonly observed—for example for particle sizes, tolerance to drugs, loudness of sounds, lifetimes of bearings, and incomes.[10,11] They are characterized by the geometric mean, μ_g, which in this application is the antilog of the mean of the logarithms of the concentrations, and by the geometric standard deviation, σ_g, which is the antilog of the standard deviation of the logarithms of concentrations. The antilog of a number N is simply 10^N. These parameters may be calculated readily from the concentration data converted into logarithmic form. If there are many zero values (which cannot be converted), the data may be converted to the cumulative fraction of values equal to or less than the stated concentration, and the methods given in the next section may be applied.

Lognormal Model for Dose-Response Curves

A similar model may be used to derive the dose-response curve. In this derivation response, R means the fraction of a large exposed population showing an adverse effect. Its value ranges from 0 to 1. It does not mean the magnitude of an effect on one individual. As the levels of the toxic doses are raised, at first a few very sensitive individuals will succumb, then the bulk of them will succumb, and finally the last few highly resistant individuals will succumb. As above, the level at which each individual

succumbs, L_i, may be equated to the median level, L_{50}, multiplied by a series of N random factors, Mk_i, each relative to its median value:

$$L_i = L_{50}\, M1_i\, M2_i\, M3_i \ldots\ldots MN_i \qquad (3)$$

In this case therefore, $M1_i$ may be a function of the individual's weight divided by the same function of the median weight of the individuals in the exposed population, $M2_i$ a relative function of its age; $M3_i$ of its sex; $M4_i$ of its state of health; $M5_i$ of its activity level; and $M6_i$ ... MN_i, of its nutrition factors, genetic factors, measurement errors, and so on. As in the previous derivation, each of these relative factors may differ randomly for each observed individual. As in eq 2, the logarithms of the observed L_i values can be represented as a constant plus the sums of N random variables. Although some small special populations, such as a select group of pure-bred animals, may follow a special pattern, it is likely that for a large varied population there will be many random modifying multiplicative factors (a high value of N). Therefore, the values of $\log(L_i)$ tend to follow a normal distribution, and the values of L_i follow a lognormal distribution. Because the response is defined as the fraction of the exposed population succumbing to the adverse effect of all doses up to the specified level, the dose-response curve must approach a cumulative lognormal distribution.

The parameters of this distribution are the median level of dose, at which 50% of the exposed population suffers the adverse effect, D_{50}, and the geometric standard deviation of the response, σ_R. The latter is a measure of the biological diversity of resistance of the members of the population to the toxic agent. For this population, similar values might be estimated for similar toxic agents. The parameters can be estimated from the data in the region of interest at the low dose end of the left tail of the distribution and applied to this region, even if the data deviate at higher doses.

If the data are available for exposures of the population to constant concentrations, the following procedure may be used for this estimation: Each dose, D_i, and associated fraction of the population responding, R_i, is converted into a derived pair of values, $\log(D_i)$ and $Z(R_i)$. The $Z(R_i)$ values are obtained as values of the standard normal distribution function, using R_i to represent the fraction of the area under the normal distribution curve. Values of Z increase with R, but are zero at $R_i = 0.5$, and negative below that value of R_i. If a table lists only positive values, the value listed for $(1 - R_i)$ may be used, and the sign of the calculated Z made negative. The following equation then may be applied to the derived pairs of data values to obtain the parameters:

$$\log(D_i) = \log(\sigma_R)\, Z(R_i) + \log(D_{50}) \qquad (4)$$

A plot of $\log(D_i)$ values as the ordinate against the $Z(R_i)$ values as the abscissa will be a straight line and will have a slope of $\log(\sigma_R)$ and an intercept of $\log(D_{50})$. The intercept occurs at values of $\log(D_i) = 0$, or $D_i = 1$. The straight line may be fitted by the usual method of least squares. If D_i data for only two values of R_i, are obtained or estimated, the parameters may be calculated by solving two simultaneous equations in the form of eq 4. For example, if dose values are obtained or estimated for 1% and 5% response values, the corresponding values of Z, -2.326 and -1.645, may be entered.

A rapid estimate of the parameters may be obtained by plotting the pairs of D and R values on lognormal probability paper and fitting a straight line to the data by eye. On this line the value of D (on the logarithmic axis) at the 50% value of R is equal to D_{50}. The value of Z(0.1587) is equal to -1. Thus, the value of σ_R is equal to $D_{50}/D_{15.87}$. This may be shown by setting up two equations as above using the value of D_{50} in the first, and of $D_{15.87}$ in the second and entering the corresponding values of Z. The second equation is subtracted from the first, and the result is solved for the value of σ_R.

If the available data are for the population being exposed to fluctuating concentrations, a more complicated procedure may be used to estimate the parameters (as described in the appendix).

Model for Integrating Risks during a Time Period

Monitoring data for fluctuating concentrations are usually recorded as values averaged for constant time periods, such as five minutes or one hour. These values also conveniently may be used as doses for the dose response

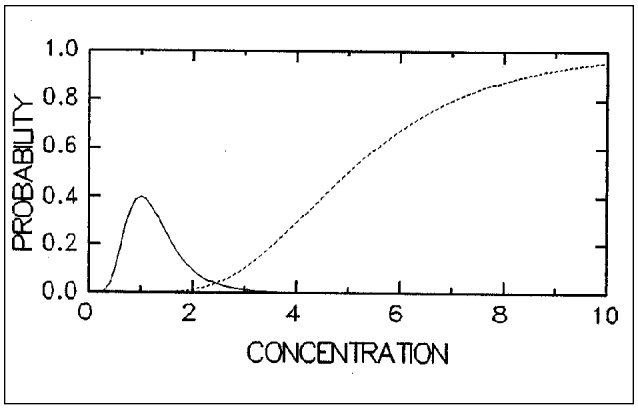

Figure 1. Solid curve: Illustrative probability density (lognormal distribution) of time-averaged concentrations that may be experienced by an individual. Parameters are $\mu_g = 1$, $\sigma_g = 1.5$. Dashed curve: Illustration of a dose-response curve (cumulative lognormal distribution) showing the fraction of population adversely affected. Parameters are $C_{50} = 5$, $\sigma_R = 1.5$. The overlapping area is the region producing health effects.

Table 1. Values of safety factor exponent, E.

	Log(P_{CR})												
Variability Ratio	**-12**	**-11**	**-10**	**-9**	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1.301**	**-1**
0	7.035	6.707	6.362	5.999	5.613	5.201	4.755	4.265	3.719	3.090	2.326	1.645	1.282
0.5	7.863	7.493	7.114	6.707	6.276	5.814	5.315	4.768	4.158	3.455	2.601	1.839	1.433
0.75	8.790	8.381	7.951	7.498	7.017	6.502	5.942	5.331	4.649	3.863	2.908	2.056	1.602
1	9.948	9.480	8.990	8.476	7.934	7.355	6.722	6.031	5.259	4.370	3.290	2.326	1.813
1.5	12.676	12.092	11.482	10.823	10.103	9.345	8.569	7.689	6.705	5.571	4.194	2.965	2.310
2	15.743	14.966	14.204	13.441	12.593	11.590	10.629	9.537	8.316	6.910	5.202	3.678	2.866
2.5	19.021	18.054	17.048	16.131	15.199	14.036	12.799	11.484	10.014	8.321	6.264	4.429	3.451
3	22.374	21.276	20.007	18.866	17.835	16.580	15.032	13.487	11.761	9.772	7.357	5.202	4.053
3.5	25.748	24.579	23.029	21.654	20.477	19.173	17.302	15.524	13.537	11.249	8.468	5.987	4.665
4	29.127	27.929	26.064	24.512	23.118	21.788	19.597	17.585	15.334	12.741	9.592	6.782	5.284
4.5	32.505	31.297	29.097	27.442	25.755	24.411	21.909	19.660	17.144	14.245	10.724	7.582	5.908
5	35.882	34.671	32.119	30.419	28.388	27.035	24.238	21.747	18.963	15.757	11.862	8.387	6.535

curve. Because D_{50} has units of concentration x time, C_{50} is defined as equal to D_{50} divided by the concentration-averaging time. The term C_{50} as used in this model does not represent μ_g, the geometric mean of the concentration pattern. The concentration-averaging time should be chosen at an appropriate value for both the measurement of concentrations and for the dose-response curve. Both the lognormal concentration distribution and the cumulative lognormal dose-response curve in a typical calculation then may be illustrated as in Figure 1. The horizontal axis represents time-averaged concentrations and the vertical axis represents probability. In the case of concentrations, it is the probability density of a single concentration during one concentration-averaging time period, and in the case of the dose-response curve it is the cumulative probability, or fraction of the population responding adversely to an exposure to the same concentration for one concentration-averaging time period.

The majority of health risks occur in the concentration range of overlap of the right tail of the left curve and the left tail of the right curve. The joint probability of a response, P_{CR}, for the plotted concentration data may be calculated numerically by computer by integrating the products of the two ordinates, P_C and R_C, over this concentration range. This joint probability represents the total risk for one concentration-averaging time period. It can be anticipated in two limiting cases. For a constant concentration, the geometric standard deviation of the left curve is 1, and the left curve condenses to a single vertical line of unit area. The joint probability in this case is therefore equal to R_C, the ordinate of the dose-response curve at this concentration. If the biological variability of response approaches 0, then the dose-response curve becomes a straight vertical line from 0 to 1 at the C_{50} value, and the joint probability becomes the area under the concentration curve to the right of this value. The joint probability is a function of the two parameters of each curve.

It is mathematically convenient to convert the four parameters into two generalized parameters. Then the joint probability of a response can be more conveniently expressed graphically or in tables in terms of only two variables and, therefore, can be determined by interpolation without the use of a computer. The two generalized parameters are defined by the author as follows:

$$V = \log(\sigma_g) / \log(\sigma_R) \quad (5)$$

$$E = \log(C_{50} / \mu_g) / \log(\sigma_R) \quad (6)$$

$$\text{or } \mu_g = C_{50} / (\sigma_R)^E \quad (6a)$$

V is the variability ratio of the concentration distribution to that of the dose-response curve. It may vary from 0 to as high as 5. E is the exponent of the safety factor. The safety factor is defined as the ratio of C_{50} to the allowable μ_g for a given allowable probability of an adverse effect. V and E can be calculated from the parameters of the two distributions.

From these two generalized parameters, the joint probability, P_{CR}, may be obtained by interpolation of the V and E values using Table 1 or Figure 2. Alternatively, for a given value of V and an allowable P_{CR}, the value of E may be obtained, and then the value of the allowable μ_g may be calculated from eq 6a. Here the denominator of the term on the right is the safety factor. For each successive averaging-time interval, the additional fraction of the population suffering the adverse effect is equal to P_{CR}

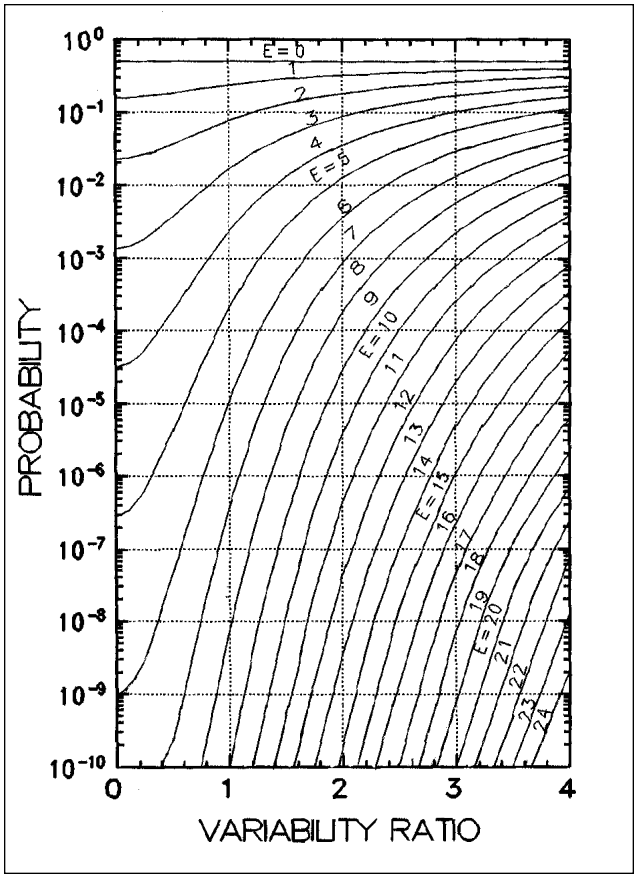

Figure 2. Relationship between the probability, P_{CR}, of an effect for a single time-averaging period, the variability ratio, V (defined in eq 5), and the safety factor exponent, E (defined in eq 6). The probability of an effect for n time-averaging periods is n P_{CR}, as shown in eq 7.

multiplied by the unaffected fraction of population remaining from the previous intervals.

If the first averaging-time interval produces a very small value of P_{CR}, the unaffected fraction will remain very close to 1 and the vulnerability pattern of this fraction of the population likely will remain substantially unchanged. Almost all of the adverse effect is produced by the high concentrations, which have a low probability. Thus, many intervals may occur before such a concentration actually occurs. For example, if $P_n = 10^{-4}$ after n averaging-time intervals, not a single adverse effect may have occurred in a population of one thousand. It may be assumed that the individuals that do not succumb recover from the effects before the next high concentration occurs. Furthermore, many of the individuals may succumb—not because of a fixed vulnerability—but because of temporary random modifying factors such as an injury, disease, or accidental presence in a high concentration hot spot. For such a population it may be assumed that the dose-response curve and, consequently, the value of P_{CR} remain constant for successive averaging-time intervals. Therefore, when the health risk rates are small, the values of P_{CR} are simply additive when calculating the total risk for a number of successive intervals. Thus, if the desired time for expressing the risk rate is n time periods, the rate is:

$$P_n = n\, P_{CR} \tag{7}$$

This equation is accurate for values of P_n up to 0.05. It yields a safe upper limit of P_n if the dose-response curve actually is changing during successive intervals because of depletion of sensitive individuals.

If the parameters of the concentration distribution (or of the dose-response curve) are changing with time, then the values of P_{CR} may be calculated at appropriate averaging-time intervals. Instead of using eq 7, the value of P_n may be calculated as the integral of P_{CR} over the total number of time intervals. If P_n is expected to exceed 0.05, the calculation can be made with allowance for the declining unaffected fraction of the population as follows:

$$P_n = 1 - \prod_{i=1}^{n} \left(1 - P_{CR}\right)_i \tag{7a}$$

The mathematical model can be adapted to evaluate effects that are believed to require two hits. If two hits are required to produce an effect during this period, then the joint probability, PP_n, during time interval n must include a calculation for each probability. If f is the fraction of the time intervals before the first hit of substance F, and (1-f) is the remaining fraction of time intervals for the second hit by substance S (which could be the same as F), then:

$$PP_{fn} = f\, n\, (P_{CR})_F\, (1\text{-}f)\, n\, (P_{CR})_S = (f - f^2)\, n^2\, (P_{CR})_F\, (P_{CR})_S \tag{8}$$

The fraction function is 0 at f values of 0 and 1, and reaches a parabolic peak of 1/4 at an f value of 1/2. It may be integrated for all possible values between 0 and 1, giving the final result:

$$PP_n = (1/6)\, n^2\, (P_{CR})_F\, (P_{CR})_S \tag{9}$$

This assumes that PP_n is not affected by the time interval between the F and S hits, or by the possibility of a second prior F hit. If the adverse effect occurs when the S hit is first, as well as when the F hit is first, then the value of the constant 1/6 is doubled to 1/3 in eq 9. It may be difficult to get the dose-response data for the two substances, especially if no symptoms appear for exposures to each substance alone. For experimental data obtained at constant concentrations, the value of each P_{CR} is simply the ordinate of its dose-response curve at the corresponding concentration, and eq 9 may be solved for the joint product of the two ordinates. If the two substances are the same and the two dose-response curves are assumed to be identical, then the ordinate of each is simply the square root of this product.

Illustrative Calculations

The use of these equations may be illustrated by the following hypothetical examples. Because the concentration averaging time affects the values of C_{50} and σ_g, it should always be specified for calculated concentrations.

Example 1. The hourly-averaged concentration levels of a toxic substance affecting 1% and 5% of the population are 1.95 and 2.57 ppm (1-hr), respectively. Calculate the parameters of the dose-response curve.

Answer: Entering these values and the values of Z of -2.326 and -1.645, respectively, yields the following two equations derived from eq 4:

$$\log(1.95) = \log(\sigma_R)\,(-2.326) + \log(C_{50})$$
$$\log(2.57) = \log(\sigma_R)\,(-1.645) + \log(C_{50})$$

Subtracting the first equation from the second and entering the log of 2.57/1.95 yields an equation without the C_{50} term that can be solved for the value of σ_R. Either equation can then be solved for C_{50}. The resulting values are σ_R = 1.50 and C_{50} = 5.00 ppm (1-hr). These are the parameters of the dose-response curve illustrated in Figure 1.

Example 2. Assume the previous dose-response parameters and that the geometric standard deviation, σ_g, of the concentrations is 2.25. Calculate the allowable geometric mean concentrations that will ensure that the affected population will not exceed 0.1%, 1%, and 5% per 40-hour week.

Answer: From eq 5, V = log(2.25) / log(1.5) = 2.0. From eq 7, we divide the P_n values of percentages by 40 to give values of P_{CR} of 2.5 x 10^{-5}, 2.5 x 10^{-4}, and 1.25 x 10^{-3}. From Figure 2, at these values for the variability ratio of 2.0, the E values are 9.1, 7.8, and 6.7. From eq 6a, the allowable geometric mean concentrations are:

For P_{40} = 0.001, $\mu_g = 5 / (1.5)^{9.1}$ = 0.12 ppm (1-hr)
For P_{40} = 0.01, $\mu_g = 5 / (1.5)^{7.8}$ = 0.21 ppm (1-hr)
For P_{40} = 0.05, $\mu_g = 5 / (1.5)^{6.7}$ = 0.33 ppm (1-hr)

Table 1 may be used for more precise calculations, using logarithmic interpolation of the probability values as follows: The logarithm of 2.5 x 10^{-5} is 0.3979 - 5. On the horizontal line for V = 2.0, this falls between the values 9.537 and 8.316. Thus, E = 9.537 - 0.3979 (9.537 - 8.316) = 9.051. This replaces the 9.1 in the line above for P_{40} = 0.001, giving a value of 0.127 ppm (1-hr). Similar calculations give, for P_{40} = 0.01, the values E = 7.756 and μ_g = 0.215 ppm (1-hr); and for P_{40} = 0.05, E = 6.744 and μ_g = 0.325 ppm (1-hr).

Example 3. Assume the same parameters as in Example 1 for the dose-response curve, and values for the concentration distribution of σ_g = 1.5 and μ_g = 0.215 ppm (1-hr). What is the risk for the total time of 1 year (8760 hr)?

Answer: From eq 5, the variability ratio V = 1.0. From eq 6:

$$E = \log(5.0 / 0.215) / \log(1.5) = 1.367 / 0.1761 = 7.760$$

From this pair of V and E values, the interpolation of Figure 2 gives P_{CR} = 2.3 x 10^{-8}. Multiplying this by 8760 gives P_{8760} = 2.0 x 10^{-4}. If Table 1 is used for more accurate calculations, on the line for V = 1.0, the value of 7.760 falls between 7.934 for $P_{CR} = 10^{-8}$ and 7.355 for $P_{CR} = 10^{-7}$. Thus, $\log(P_{CR})$ = [(7.934 - 7.760) / (7.934 - 7.355)] - 8 = 0.301 - 8, and P_{CR} = 1.998 x 10^{-8}. Multiplying this by 8760 gives P_{8760} = 1.75 x 10^{-4}, a more precise value. Comparison of this P_{8760} = 2 x 10^{-4} with the P_{40} = 0.01 for the second item in example 2, which has the same μ_g, shows the large effect of the value of σ_g.

DISCUSSION

Dose more accurately refers to concentrations inside the body, rather than to external ones. In toxicological studies at constant concentrations, the internal concentrations are not far from equilibrium with the external ones and, therefore, no serious inaccuracies arise from using the latter. However, when concentrations fluctuate, as they usually do, the internal concentration patterns may differ substantially from the external ones, depending upon various toxicokinetic factors. This problem was studied experimentally for carbon monoxide in a previous paper.[12] A practical method was then developed[13,14] for using the biological half-life of a pollutant to convert external concentrations to "biologically effective concentrations." These are values at in-vivo equilibrium with the internal concentrations on the basis of pollutant absorption and excretion or destruction in a one-compartment kinetic model. Practical methods for a two-compartment kinetic model also were developed to calculate a "biological damage parameter," which provides for the kinetic process of damage repair as well. Where appropriate, it is believed that use of one of these two parameters would be more accurate than use of the external concentrations for risk rate calculations.

It may be seen from Figure 1 that most of the risk results from a small portion of the concentration distribution at its extreme right tail. Therefore, an accurate fit of the theoretical equation to the data in this region is important. Atmospheric aerosols have been known to follow a bimodal lognormal distribution, resulting from two different simultaneous processes. In a study[15] that sampled trace metal concentrations in the air at 59 sites

in 8 cities for one year each, histograms of the concentrations at each site frequently showed a small second peak at high concentrations. It is believed that this may be allowed for by the following procedure. The frequency for each concentration interval in the region of the second peak, P_C, calculated from the distribution parameters in eq 10 (see Appendix), is deducted from the observed frequency to give the excess frequency. The sum of these excesses, each multiplied by the interval width, is the excess area. Each excess frequency x width is divided by the excess area to become one fractional value of a new lognormal distribution. The parameters of the latter are calculated. The P_{CR} calculated from this new distribution may be weighted by the fractional excess area and then added to the original P_{CR} weighted by 1 minus the fractional excess area.

In some populations there may be a small group of exceptionally vulnerable individuals. If there is a bimodal distribution of vulnerabilities, a different dose-response curve may be derived for each group. A separate calculation of probability may be made for each group. The probabilities then may be added after being weighted according to the fraction of population in each group.

Assessment of very low health risk rates requires extrapolations of models to far below any possibility of experimental validation. Many methods have been based upon arbitrary assumptions of models and safety factors. Extreme upper limits of estimates have been used for assurance of safety. The accuracy of statistical calculations (including those of this proposed method) at the extreme limits of an assumed normal distribution depends upon the presence of a very large number (previously referred to as N) of random variables affecting the observed variable. The method proposed here has the advantages of being based upon scientific principles rather than arbitrary safety factors and of allowing the types of calculations illustrated, which are not otherwise possible. The concentration and dose-response data may be compactly described, each by two parameters, and the calculations are simple to make. An estimate of the central value is determined. Limits may be obtained by selecting appropriate ranges of the four parameters and repeating the calculations.

Several important aspects are clarified. In addition to the geometric mean concentration, the geometric standard deviation also has a major influence upon the degree of health risk. It is also evident that health risk rates can never be reduced to 0. While this fact may be emotionally disturbing to some, it is more realistic than the "safe" categorization for levels below the Threshold Limit Value or Permissable Exposure Limit values. It is hoped that the availability of this theoretical model will encourage investigators to accumulate the necessary data to determine the needed parameters. Toxicological data for small, select pure-bred animal groups may not be a good source for estimates of the geometric standard deviation of the dose-response curve for large diverse human populations, which is just as important as the C_{50} parameter. If models other than those proposed better fit the data, they may be used for a similar type of calculation.

This method makes possible a new type of environmental standard that specifies the two parameters of the dose-response curve for humans for a particular pollutant, and the maximum allowable health risk rate. The permissible geometric mean concentration of a pollutant will be higher if the geometric standard deviation is lower. More accurate evaluations of health risk rates may reduce the need for large safety factors and, thereby, the cost of needed controls.

APPENDIX

The derivations of the equations used above and the nomenclature are given below. Although these equations may appear complex, the actual working equations used in the calculations are not, as has been shown in the illustrative calculations.

Derivations of Equations

The theoretical basis for the generalized parameters—variability ratio V and safety factor exponent E, defined in eqs 5 and 6 respectively—and for the method of calculating P_{CR} may be developed as follows: The lognormal pattern defines the probability densities at each concentration as:

$$P_C dz = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \qquad (10)$$

$$\text{where } z = \frac{\log (C/\mu_g)}{\log (\sigma_g)} \qquad (11)$$

The cumulative lognormal pattern for the dose-response curve is defined as:

$$R_C = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} e^{-x^2/2}\, dz \qquad (12)$$

$$\text{where } Z = \frac{\log (C/C_{50})}{\log (\sigma_R)} \qquad (13)$$

In eq 12, x is a dummy variable that disappears after the integration. Equation 4 was derived by rearranging eq 13

after replacing C/C_{50} with D/D_{50} and R_C with R_D. Equation 13 also may be rearranged as follows:

$$Z = \frac{\log (C/\mu_g)}{\log (\sigma_R)} - \frac{\log (C_{50}/\mu_g)}{\log (\sigma_R)} \quad (14)$$

Combining eqs 5, 6, 11, and 14 and rearranging yields:

$$Z = V z - E \quad (15)$$

The joint probability of the overlapped area between the two curves is:

$$P_{CR} = \int P_C R_C dz = \frac{1}{\sqrt{2\pi}} \int_{-12}^{12} e^{-z^2/2} \left[\frac{1}{\sqrt{2\pi}} \int_{-12}^{Vz-E} e^{-x^2/2} dx \right] dz \quad (16)$$

The integral in the bracketed area on the right is the cumulative lognormal distribution representing the value of R for each value of z. This integral may be evaluated by entering the upper limit of integration in the normal distribution function available in many mathematical computer programs. The entire integral then may be evaluated in terms only of the two given values of V and E, instead of in terms of the four parameters of the two overlapping curves. Theoretically, the lower limits of the two integrations are minus infinity. It was found that -12 was sufficiently low to give accurate results even down to probabilities of 10^{-12}. The upper limit of the left integration was set at 12 rather than at plus infinity. The numerical integration was made using the Mathcad® program. Results are given in Table 1 and Figure 2.

Determining the Parameters of the Dose-Response Curve

The parameters of the dose-response curve may readily be determined if data are available for population exposures to constant concentrations, by using eq 4 as described previously. If the exposures are to fluctuating concentrations, the calculation is more complex. The following equation may be used as an empirical approximation to the data in Table 1 and Figure 2:

$$E = 0.73 (V + 0.25 + 0.21 / V) \log(9.0 / P_{CR}) \quad (17)$$

The root mean square deviation of this E from the correct value was 0.2 in the ranges of V = 0.5 to 4, and P_{CR} = 0.0001 to 0.05. From this equation and eqs 5 and 6, the following equation may be derived for the calculation:

$$A = U_i + X_i B + Y_i B^2 \quad (18)$$

It is necessary to have two or more sets of data, each describing a concentration pattern by its μ_g and σ_g parameters, and its associated P_{CR} and all within the accurate ranges cited above for eq 17. From the values for each set the three functions U_i, X_i, and Y_i are calculated, as defined below. Then the values of the two constants, A and B, may be found that best fit all the data sets, each expressed in the form of eq 18. A and B are the values of $\log(C_{50})$ and $\log(\sigma_R)$, respectively. In this equation U_i, X_i, and Y_i are calculated as follows:

$$U_i = \log(\mu_g)_i + \log(\sigma_g)_i [0.70 - 0.73 \log(P_{CR})_i] \quad (19)$$

$$X_i = 0.174 - 0.182 \log(P_{CR})_i \quad (20)$$

$$Y_i = [0.146 - 0.153 \log(P_{CR})_i] / \log(\sigma_g)_i \quad (21)$$

After the approximate values of C_{50} and σ_R are obtained, more precise values may be derived by using a simplex procedure[12] for adjusting them in small steps and then applying the more precise values in Table 1 until the best fit to the data is obtained.

Nomenclature

- A $\log(C_{50})$
- B $\log(\sigma_R)$
- C concentration
- C_i individual time-weighted-average concentration
- C_m median concentration
- C_{50} concentration equal to D_{50} divided by the concentration-averaging time
- D_i individual dose
- D_{50} dose causing an adverse response in 50% of the population
- E safety factor exponent, defined in equation 6
- f fraction of time intervals until first hit
- F substance required for first hit
- L_i level of dose toxic to one individual
- L_{50} median level of toxic dose for a population
- Mj_i random individual multiplicative factor caused by item j , relative to its median value
- Mk_i random individual multiplicative factor caused by item k, relative to its median value
- n number of concentration-averaging time periods
- N a number
- P_C frequency of concentration C for one concentration-averaging time period
- P_{CR} joint probability of the concentrations and responses from the overlapping curves

- P_n risk of an adverse effect within n concentration-averaging time periods
- PP_{fn} risk of an adverse effect requiring two hits, with the first hit at f fraction of n total periods
- PP_n risk of an adverse effect requiring two hits within n concentration-averaging time periods
- R_i response of an exposed population to one concentration pattern
- R_C ordinate of the dose-response curve at concentration C
- S substance required for second hit
- U_i data function defined by eq 19
- V variability ratio of the parameter for concentrations to that of the dose-response curve, defined by eq 5
- x dummy variable for integration
- X_i data function defined by eq 20
- Y_i data function defined by eq 21
- z variable for standard normal distribution derived from concentration data
- Z variable for standard normal distribution derived from dose-response data
- $Z(R_i)$ variable for standard normal distribution derived from response R_i
- μ_g geometric mean concentration
- σ_g geometric standard deviation of the concentration pattern
- σ_R geometric standard deviation of the dose-response curve

ACKNOWLEDGMENT

Supported in part by NIOSH Grant 5 T 42 OH 07091.

REFERENCES

1. *Issues in Risk Assessment;* Committee on Risk Assessment Methodology, National Research Council, National Academy Press: Washington, DC, 1993
2. *Researching Health Risks;* Office of Technology Assessment, U.S. Congress, U.S. Government Printing Office: Washington, DC, 1993; OTA-BBS-570.
3. *The Risk Assessment of Environmental and Human Health Hazards: A Textbook of Case Studies;* Paustenbach, D.J., Ed; John Wiley & Sons: New York, 1989.
4. Ricci, P.F.; Molton, L.S. *Environ. Sci. Technol.* **1985**, *19*, 473-479.
5. Food Safety Council. *Food and Cosmetics Toxicol.* **1978**, *16* (Supplement), 1-136.
6. Munroe, I.C.; Krewski, D.I. *Food and Cosmetics Toxicol.* **1981**, *19*, 549-560.
7. *Approaches to Risk Assessment for Multiple Chemical Exposures*; Stara, J.F.; Erdreich, L.S., Eds.; Office of Research and Development, Environmental Protection Agency: Cincinnati, OH, 1984; EPA 600/9-84-008.
8. Hattis, D. *Strategies for Assessing Human Variability in Susceptibility, and Using Variability to Infer Human Risks*; Presented at International Life Sciences Institute Conference on Interindividual Variability; ILSI Press: Washington, DC, 1997.
9. Rees, D.C.; Hattis, D. In *Principles and Methods of Toxicology*, 3rd ed. Hayes, A.W., Ed.; Raven Press: New York, 1994, Chapter 8.
10. Aitcheson, J.; Brown, J.A.C. *The Lognormal Distribution*; University Press: Cambridge, 1957.
11. *Lognormal Distributions. Theory and Applications;* Crow, E.L.; Shimizu, K, Eds.; Marcel Dekker: New York, 1988.
12. Saltzman, B.E.; Fox, S.H. *Environ. Sci. Technol.* **1986**, *20,* 916-923.
13. Saltzman, B.E. *Am. Ind. Hyg. Assoc. J.* **1988**, *49*, 213-225.
14. Saltzman, B. E. *J. Air & Waste Manage. Assoc.* **1996**, *46*, 1022-1034.
15. Saltzman, B.E.; Cholak, J.; Schafer, L.J.; Yeager, D.W.; Meiners, B.G.; Svetlik, J. *Environ. Sci. Technol.* **1985**, *19*, 328-333.

About the Author

Bernard E. Saltzman is an Emeritus Professor of Environmental Health in the College of Medicine, Department of Environmental Health, University of Cincinnati, Cincinnati, OH 45267-0056 (or send e-mail to Bernard.Saltzman@uc.edu). He previously retired from the U.S. Public Health Service after working in their early air pollution and industrial hygiene programs.