A CONCEPTUAL MODEL FOR GENERATING RANDOM EXPOSURES FOR USE IN COMPUTER SIMULATIONS

Paul Hewett Ph.D. CIH

A CONCEPTUAL MODEL FOR GENERATING RANDOM EXPOSURES FOR USE IN COMPUTER SIMULATIONS

Paul Hewett Ph.D. CIH

1 ABSTRACT

Computer simulation is usually the only means for predicting the ability of a corporate sampling strategy to detect a poorly controlled work environment. To date, several papers have been published that use computer simulation to evaluate particular sampling strategies, but information on how to generate random exposures from a hypothetical exposure group is not readily available. A conceptual model is presented that can be applied to either conventional, single-shift TWA exposure limits (L_{TWA}) or the less common long-term average exposure limits (L_{LTA}). Using this model, procedures for generating random exposure values are presented.

In addition, equations are derived from the model for calculating the 95th percentile "worker 95th percentile exposure" and the 95th percentile "worker mean exposure". These equations can easily be modified to calculate other percentiles. Equations are also presented for calculating the fraction of workers (θ_P) expected to have an individual 95th percentile exposure greater than the L_{TWA}, and the fraction of workers (θ_M) expected to have an individual mean exposure greater than the L_{LTA}. Such calculations can be used in an educational setting to estimate and visualize (via graphs) the distributions of worker 95th percentiles and worker mean exposures. In summary, the model presented can be used to (a) design exposure assessment strategies through computer simulation, and (b) train industrial hygienists.

2 INTRODUCTION

A recurring theme among many research papers describing exposures and the workplace determinants of exposures is that exposure variability can be divided into within-worker and between-worker components. The within-worker component reflects the fluctuations in individual exposures due to the day-to-day variations in control efficacy, production levels, work practices, and the assigned tasks and activities. The between-worker component reflects systematic differences between workers, due to consistently different work practices and controls, for example. To date, the use of this components-of-variance (COV) model has been restricted to the situation where there is a true long-term average exposure limit (L_{LTA}) (Rappaport et al., 1995; Lyles et al, 1997a; Hewett, 2001). The primary objectives of this paper are to (a) apply this COV model to the far more common single shift, TWA exposure limits (L_{TWA}), (b) show how it can be used to generate random exposures for computer simulations, and (c) illustrate several uses of the model.

3 VARIABLES

L	- exposure limit
L_{TWA}	- exposure limit for single-shift, time-weighted average exposures
L_{LTA}	- exposure limit for long-term average exposures
x	- a random, full-shift exposure for a randomly selected worker in a specific exposure group
G	- geometric mean for the group exposure profile
G_P	- geometric mean for the statistical distribution of "worker 95th percentiles"
G_M	- geometric mean for the statistical distribution of "worker means"
G_k	- geometric mean exposure for the kth worker
D	- geometric standard deviation for the group exposure profile
D_b	- between-worker geometric standard deviation
D_w	- within-worker geometric standard deviation

ρ - rho; group heterogeneity coefficient

X_q - the q·100%-tile of the group exposure profile; e.g., $X_{0.95}$ = 95^{th} percentile of the group exposure profile
P_q - the q·100%-tile of the statistical distribution of "worker 95^{th} percentile exposures"; e.g., $P_{0.95}$ = 95^{th} percentile of the distribution of "worker 95^{th} percentiles"
P_k - the upper percentile exposure for the *kth* worker; e.g., the 95^{th} percentile exposure for the *kth* worker
M_q - the q·100%-tile of the statistical distribution of "worker mean exposures"; e.g., $M_{0.95}$ = 95^{th} percentile of the distribution of "worker means"
M_k - the mean exposure for the *kth* worker; e.g., the mean exposure for the *kth* worker
$\bar{M}$ - arithmetic mean for the group exposure profile; mean of the "worker mean exposures"

Z_x - Z-value corresponding to a percentile of the group exposure profile
Z_P - Z-value corresponding to a percentile of the statistical distribution of "worker 95^{th} percentile exposures"
Z_M - Z-value corresponding to a percentile of the statistical distribution of "worker mean exposures"
K - Z value that determines the distribution of worker upper percentiles; e.g., K=1.645 when we are interested in the distribution of "worker 95^{th} percentile exposures"

θ - theta; fraction of the group exposure profile that exceeds an exposure limit L; $P(x > L)$
θ_P - fraction of the distribution of individual worker upper percentiles that exceed the L_{TWA}; $P(P_k > L_{TWA})$
θ_M - fraction of the distribution of individual worker mean exposures that exceed the L_{LTA}; $P(M_k > L_{LTA})$

4 BACKGROUND

Oldham and Roach (1952) were the first to describe the use of the COV model for analyzing exposure data. In their study of coal workers in the United Kingdom repeat measurements (short-term measurements expressed in million particles per cubic foot) were collected from randomly selected workers from a number of coal mines. They may also have been the first to use the lognormal distribution to describe occupational exposure data. In the U.S., the COV model was not routinely used until the 1980's. Petersen et al. (1986) used COV analysis of a repeat measurement dataset to help design a sampling strategy for an industry-wide epidemiology study. Spear et al. (1987) used the technique for analyzing complex datasets representing a variety of plants where a selection of workers were each monitored more than once.

Numerous investigators have since used the COV model to evaluate and quantify the within- and between-worker sources of exposure variability. Spear and Selvin (1989) (see also Spear, 1991) discussed compliance with regulatory limits in the context of the COV model applied to occupational exposure data.\[a] A particular group of researchers (Rappaport et al, 1995; Lyles et al, 1997a; Tornero-Velez et al, 1997b) have gone a step further and recommended that it be used for determining compliance whenever there is a true long-term average exposure limit.\[b] In this paper, it is shown how this COV model can be extended to the more common single shift limits, or L_{TWA}. Recently, this COV model was used to generate random exposures for use in determining the power (i.e., ability) of a particular sampling strategy to detect poorly controlled work environments, relative to a L_{TWA} (Hewett, 1999).

\[a] Aitchison and Brown (1957) had earlier developed a similar COV model which they applied to the distribution of incomes within society.

\[b] Until recently, the only official long-term average exposure limit for a gas, vapor, or particulate was the 3 ppm annual mean limit for vinyl chloride used by several European Union nations. For example, the United Kingdom had a dual limit: a 3 ppm annual mean limit, as well as a conventional single shift limit of 7 ppm. Single shift exposures may exceed the long-term limit, but not the single shift limit. The annual average should not exceed the long-term limit. This long-term average exposure limit was recently rescinded and replaced with a single shift, TWA limit of 3 ppm (HSC, 2002).

5 THE COMPONENTS-OF-VARIANCE MODEL

Let us assume that each worker experiences roughly 250 single shift exposures per year. The number of exposures is sufficiently large to approximate a continuous distribution, and experience has shown that individual worker exposures to gases, vapors, and particulates are well described by the lognormal distribution. The COV model, as applied to lognormally distributed exposures, is completely described by the following relation:

$$x \sim L(G, D, \rho)$$

which translates as x, a random exposure from a randomly selected worker, is lognormally distributed with a "group" geometric mean (G), a "group" geometric standard deviation (D), and a group heterogeneity coefficient (ρ). Appendix A contains a formal derivation of this model.

The group geometric standard deviation has two components: the between-worker and within-worker geometric standard deviations. We will use D_b to refer to the geometric standard deviation for the statistical distribution of "worker geometric means". Similarly, D_w refers to the within-worker geometric standard deviation for the distribution of day-to-day, full-shift exposures for a randomly selected worker. The relationship between D, D_w and D_b is as follows:

$$(\ln D)^2 = (\ln D_b)^2 + (\ln D_w)^2$$

Let us define ρ as the ratio of the between-worker variability to the group total variability:

$$\rho = \frac{(\ln D_b)^2}{(\ln D_b)^2 + (\ln D_w)^2} = \frac{(\ln D_b)^2}{(\ln D)^2}$$

The variable ρ ranges between 0 and 1, and will be useful as an indicator of group heterogeneity. Values of ρ near 0 indicate that the group is nearly homogeneous, meaning that the exposure profiles[\c] of individual workers do not differ greatly and are all similar to that of the entire group. Values of ρ approaching 1 indicate that the group is highly heterogeneous, meaning that the exposure profiles of individual workers are distinctly different from each other and all are considerably different from the overall group exposure profile. Later on it will be useful to express D_b or D_w solely in terms of the group D and ρ:

$$\ln D_w = \sqrt{1-\rho} \cdot \ln D \qquad \text{Eq. 1}$$

$$\ln D_b = \sqrt{\rho} \cdot \ln D \qquad \text{Eq. 2}$$

For this paper we will use the following values of ρ to indicate low, medium, and high between-worker variability:

Low - ρ = 0.05
Med - ρ = 0.20
High - ρ = 0.40

See Appendix B for a discussion regarding how these values were determined.

The distributions of a variety of useful variables can be derived from the model. The overall group exposure profile can be described as the distribution of random exposures having a geometric mean and a geometric standard deviation:

$$\ln(x) \sim \mathbb{N}\left(\ln G, (\ln D)^2\right)$$

This relation is read as the log-transformed group exposure x is normally distributed with a mean of lnG and a variance of $(\ln D)^2$. In this model the geometric mean of the *kth* worker is lognormally distributed:

[\c] "Exposure profile" refers to the distribution of exposures for either an individual worker or an entire exposure group. For gases, vapors, and particulates, this distribution is best described using lognormal distribution parameters.

$$\ln\left(G_k\right) \sim \mathbb{N}\left(\ln G\,,\,(\ln D_b)^2\right)$$

where G_k is the geometric mean for a randomly selected *kth* worker and D_b is the between-worker geometric standard deviation. The geometric mean of the "worker geometric means" (which can be considered the between-worker geometric mean) is identical to the group geometric mean G.

The log-transformed *jth* exposure for the *kth* worker is normally distributed:

$$\ln\left(x_{jk}\right) \sim \mathbb{N}\left(\ln G_k\,,\,(\ln D_w)^2\right)$$

where the geometric mean is unique to the *kth* worker. Notice that the model requires a common variance of $(\ln D_w)^2$ for all workers. The log-transformed mean exposure for the *kth* worker can be calculated using G_k and D_w (Leidel et al., 1977):

$$\ln\left(M_k\right) = \ln G_k + \frac{1}{2}\cdot(\ln D_w)^2$$

Since the log-transformed mean, for all workers in the group, is simply shifted from the log-transformed geometric mean by a fixed amount, the distribution of log-transformed worker means is identical to that of the log-transformed worker geometric means, except shifted by this fixed quantity. Consequently, the log-transformed "worker mean" for the *kth* worker has the following distribution:

$$\ln\left(M_k\right) \sim \mathbb{N}\left(\ln G_M\,,\,(\ln D_b)^2\right) \qquad \text{where } \ln G_M=\ln G+1/2\cdot(\ln D_w)^2.$$

The log-transformed "upper percentile" exposure - e.g., the 95th percentile exposure for the *kth* worker - can be calculated using the following relationship:

$$\ln\left(P_k\right) = \ln G_k + K\cdot\ln D_w$$

where K = an appropriate Z-value; e.g., K=1.645 for the 95th percentile. Using the same reasoning applied to the distribution of worker means, the log-transformed "upper percentile" for the *kth* worker is also normally distributed (Nicas, 1992):

$$\ln\left(P_k\right) \sim \mathbb{N}\left(\ln G_P\,,\,(\ln D_b)^2\right) \qquad \text{where } \ln G_P=\ln G+K\cdot\ln D_w.$$

Since we are interested in the distribution of "worker 95th percentile exposures", let K=1.645. In this paper we will focus on individual worker 95th percentile exposures. Therefore, P_k will refer to the *kth* worker's 95th percentile exposure. Other percentiles can be considered simply by changing K in the above equation.

The probability density functions for the group exposure profile, worker geometric means, worker 95th percentile, and worker means can be calculated using the equations in Appendix C. See Figures 1 and 2 for examples of these probability density functions.\[d]

5.1 Central Tendency

The group true geometric mean (G) represents the median or 50th percentile of the group exposure profile. The true or population mean of the group exposure profile ($\bar{M}$) can be calculated from G and D using the following standard equation (Leidel et al., 1977):

$$\bar{M} = \exp\left(\ln G + \frac{1}{2}\cdot(\ln D)^2\right) = G\cdot\exp\left(\frac{1}{2}\cdot(\ln D)^2\right) \qquad \text{Eq. 3}$$

The group mean is also the mean of the "worker mean exposures".

\[d] There is no physical distribution of worker geometric means, worker 95th percentiles, or worker means. These statistical distributions exist only as useful constructs. For example, we can use the distribution of worker 95th percentiles to estimate the fraction of workers that are in compliance with a L_{TWA}.

5.2 Useful Relationships

Equations can be derived from the COV model for calculating the following:

- group upper percentile exposures; e.g., the group 95th percentile
- upper percentiles for the distribution of "worker 95th percentiles"
- upper percentiles for the distribution of "worker means"
- the group exceedance fraction (θ)
- the fraction (θ_P) of "worker 95th percentiles" expected to exceed L_{TWA}
- the fraction (θ_M) of "worker means" expected to exceed L_{LTA}

5.3 Calculation of Various Upper Percentiles

Any q·100%-tile of the group exposure profile can be calculated using the following equation:

$$X_q = \exp(\ln G + Z \cdot \ln D)$$

For example, if we are interested in the 95th percentile of the group exposure profile we replace Z with $Z_{0.95} = 1.645$:

$$X_{0.95} = \exp(\ln G + 1.645 \cdot \ln D) \qquad \text{Eq. 4}$$

Because each worker in the exposure group has a common within-worker geometric standard deviation, it follows that there will be a lognormal distribution for the 90th, 95th, 99th, or any other percentile worker exposure. The q·100%-tile of the distribution of "worker upper percentiles" can be calculated using the following equation:

$$P_q = \exp\left[\left(\ln G + K \cdot \ln D_w\right) + Z_P \cdot \ln D_b\right]$$

For example, if we are interested in the 95th percentile of the distribution of "worker 95th percentiles", we replace K with 1.645 and Z_P with 1.645. We can also substitute Equations 1 and 2 for $\ln D_w$ and $\ln D_b$ so that the final equation is in terms of G, D, and ρ:

$$P_{0.95} = \exp\left[\left(\ln G + 1.645 \cdot \sqrt{1-\rho} \cdot \ln D\right) + 1.645 \cdot \sqrt{\rho} \cdot \ln D\right] \qquad \text{Eq. 5}$$ [e]

Similarly, the q·100%-tile of the distribution of "worker means" can be calculated:

$$M_q = \exp\left[\left(\ln G + \frac{1}{2} \cdot (\ln D_w)^2\right) + Z_M \cdot \ln D_b\right]$$

For example, if we were interested calculating the 95th percentile of the distribution of "worker means", we replace Z_M with 1.645. We can also replace $\ln D_w$ and $\ln D_b$ with Equations 1 and 2 so that the final equation is in terms of G, D, and ρ:

$$M_{0.95} = \exp\left[\left(\ln G + \frac{1}{2}(1-\rho) \cdot (\ln D)^2\right) + 1.645 \cdot \sqrt{\rho} \cdot \ln D\right] \qquad \text{Eq. 6}$$ [f]

Hereafter we will assume that $P_{0.95}$ and $M_{0.95}$ refer to the 95th percentile "worker 95th percentile exposure" and 95th percentile "worker mean exposure", respectively.

[e] As ρ approaches zero, $P_{0.95}$ approaches $X_{0.95}$.

[f] As ρ approaches zero, $M_{0.95}$ approaches the group mean $\bar{M}$.

5.4 Calculation of Exceedance Fractions θ, θ_P , and θ_M

The fraction of all exposures for the group expected to exceed any particular value L (e.g., L_{TWA} , L_{LTA} , or action limit) can determined from the following relation:

$$\theta = 1 - \Phi\left[\frac{\ln L - \ln G}{\ln D}\right] \qquad \text{Eq. 7}$$

The argument of the phi (Φ) function - the quantity in the brackets - has a Z~N(0,1) distribution. The fraction of the Z distribution to the left of the argument can be obtained from any Z table found in statistics texts, or from the inverse Z function found in nearly all computer spreadsheet programs.

The fraction (θ_P) of all group workers having a 95th percentile exposure greater than a L_{TWA} can be calculated:

$$\theta_P = 1 - \Phi\left[\frac{\ln L_{TWA} - \ln G_P}{\ln D_b}\right] = 1 - \Phi\left[\frac{\ln L_{TWA} - \left(\ln G + 1.645 \cdot \ln D_w\right)}{\ln D_b}\right] \qquad \text{Eq. 8}$$

Similarly, the fraction (θ_M) of all group workers having a mean exposure greater than a L_{LTA} can be calculated:

$$\theta_M = 1 - \Phi\left[\frac{\ln L_{TWA} - \ln G_M}{\ln D_b}\right] = 1 - \Phi\left[\frac{\ln L_{LTA} - \left(\ln G + \frac{1}{2} \cdot (\ln D_w)^2\right)}{\ln D_b}\right] \qquad \text{Eq. 9}$$

---

6 GENERATION OF RANDOM EXPOSURES

When testing the ability of a sampling strategy to detect poorly-controlled group exposure profiles it is necessary to generate random exposures (Hewett, 1999, 2005). The artificial data are then analyzed and interpreted according to the specified strategy. Such artificial exposure assessment surveys are then repeated hundreds to thousands of times. Counter variables are used to keep track of the number of artificial surveys that lead to a decision that the exposure profile was acceptable. The COV model presented here can be used to generate random exposure measurements for exposure groups having specific amounts of within- and between-worker variability.

6.1 Procedure when G, D, and ρ are specified

For a fixed geometric standard deviation (D) and level of group heterogeneity (discussed in Appendix B), the corresponding group geometric mean (G) can be calculated as well as the within- and between-worker geometric standard deviations. The group geometric mean and the between-worker D are then used to computer generate a geometric mean for the exposure profile of a randomly selected worker. This worker-specific G is paired with the within-worker D to generate a random exposure. This process is repeated for as many workers (or measurements per worker) as are necessary for each survey.

For simulating sampling strategies where (a) maximum risk employees are selected, (b) one or more employees may be sampled more than once, or (c) the strategy requires the estimation of the within- and between-worker variance components, the following general steps can be used to simulate exposure measurements:

1. Specify G, D, and ρ for a hypothetical exposure group.

2. Using D and ρ , calculate D_w and D_b using Equations 1 and 2, presented earlier.

3. Generate a geometric mean (G_k) for a randomly selected worker:

$$random\ G_k = \exp\left(\ln G + Z \cdot \ln D_b\right)$$

where Z is randomly generated from a standardized normal distribution: Z ~ ℕ(0,1). Functions that generate random Z-values can be found in most programming languages, statistical packages, mathematics programs, or spreadsheets. Note that each randomly generated Z-value should be used only once.\[g]

4. Generate a random exposure for this random worker:

$$x_{jk} = \exp\left(lnG_k + Z \cdot \ln D_w\right)$$

where Z is randomly generated from a distribution that is ℕ(0,1).

6.2 Procedure when only G and D are specified

There are strategies that do not require or permit repeat sampling of each worker. The AIHA strategy (Mulhausen and Damiano, 1998) (see also Hewett, 2005) specifies that for initial surveys a single measurement should be collected from each of *n* randomly selected workers. For this and similar strategies where between-worker variability is not a factor the following general steps can be used to simulate exposure measurements:

1. Specify G and D.

2. Generate a random exposure:

$$x_j = \exp\left(lnG + Z \cdot \ln D\right)$$

where Z is randomly generated from a distribution that is ℕ(0,1).

6.3 Calculation of the Group Geometric Mean

In Sections 6.1 and 6.2 a group geometric mean G is necessary for generating random exposures. However, we often do not start with G, but with some other value derived from the group exposure profile. For example, we may be interested in determining the ability of a strategy to detect some clearly unacceptable group exceedance fraction θ; e.g., θ = 0.25 (Hewett, 2005). Consequently, we need to determine the group geometric mean (given that D has already been specified) corresponding to θ=0.25. The following equations are useful for determining the group geometric mean G that corresponds to θ, as well as other parameters of interest.

Given θ, the group exceedance fraction, calculate G :\[h]

$$Z = \Phi^{-1}\left[1-\theta\right] \qquad \text{Eq. 10}$$

$$G = \exp\left(\ln L - Z \cdot \ln D\right) \qquad \text{Eq. 11}$$

Given θ_P, the exceedance fraction for worker 95th percentiles, calculate G :

$$Z_P = \Phi^{-1}\left[1-\theta_P\right]$$

\[g] If such a function is not available, a random value from an approximate ℕ(0,1) distribution can be generated by summing twelve random values from the uniform distribution, and then subtracting six. For example, using a spreadsheet one can use the “rand” function to generate a single random uniform variate. Therefore, a random z would equal "rand+rand+...+rand - 6", where there are twelve rand functions in the equation. This is not best method for generating random z-values, but it works well enough for all but the most discriminating user. A superior method for generating random Z-values is the Marsaglia-Bray algorithm .

\[h] The function Φ^{-1} refers to the inverse of the Z-distribution, i.e., the Z-value corresponding to 1-θ.

$$G_P = \exp\left(\ln L_{TWA} - Z_P \cdot \ln D_b\right)$$

$$G = \exp\left(\ln G_P - 1.645 \cdot \ln D_w\right)$$

Given θ_M, the exceedance fraction for worker means, calculate G :

$$Z_M = \Phi^{-1}\left[1 - \theta_M\right]$$

$$G_M = \exp\left(\ln L_{LTA} - Z_M \cdot \ln D_b\right)$$

$$G = \exp\left(\ln G_M - \frac{1}{2} \cdot (\ln D_w)^2 \right)$$

Given $X_{0.95}$, the group 95th percentile exposure, calculate G :

$$G = \exp\left(\ln X_{0.95} - 1.645 \cdot \ln D\right) = \frac{X_{0.95}}{D^{1.645}} \qquad \text{Eq. 12}$$

Given $\bar{M}$, the group mean exposure, calculate G :

$$G = \frac{\bar{M}}{\exp\left(\frac{1}{2} \cdot (\ln D)^2 \right)} \qquad \text{Eq. 13}$$

7 DISCUSSION

The COV model presented by others (Lyles et al., 1997a; Tornero-Velez et al., 1997) can only be applied to true, long-term average exposure limits (Hewett, 1998a, 1998b, Leidel and Busch, 2000). Since for gases, vapors, and particulates such limits are few (Hewett, 2001), their model has limited applicability. In contrast, the expanded version presented here can be applied to the hundreds of TWA exposure limits sponsored by regulatory agencies and professional organizations, as well as true long-term average exposure limits.

7.1 Design of Exposure Assessment Strategies

Recently Hewett (1999) presented a user-friendly computer program for evaluating the ability of several commonly used exposure assessment strategies to detect poorly controlled work environments. An improved version of this program was developed for a paper on the design of performance-based exposure assessment strategies (Hewett, 2005). The COV model and equations for generating random exposures discussed here were used in this computer program. (A copy of the program and details regarding its use are available from the author and Exposure Assessment Solutions, Inc.)

The programming language available in any spreadsheet can be used to develop a computer simulation of a specific sampling strategy. The basic steps are:

- generate random exposures using a method similar to that in Section 6
- analyze the (artificial) measurements using the procedures specified by a corporate strategy, and
- interpret the results relative to a TWA exposure limit.

It is not the purpose of this paper to cover all the details of computer simulation, but in brief the process is as follows. As previously mentioned, pick a group exceedance fraction, say θ=0.25 (i.e., 25% overexposures). For a fixed group geometric standard deviation (D) and level of group heterogeneity (ρ), the corresponding group geometric mean (G) can be calculated (Equations 10 and 11), as well as the within- and between-worker geometric standard deviations (Equations 1 and 2). The group G and the between-worker geometric standard deviation (D_b) are then used to computer generate a geometric mean (G_k) for the exposure profile of a randomly selected worker. This worker-specific G_k is then paired with the within-worker geometric standard deviation (D_w) to generate a random exposure (as discussed in Section 6). This process is repeated for as many workers and measurements per worker as are necessary for each survey. The artificial data are then analyzed and interpreted according to the specified strategy. Such artificial exposure assessment surveys are then repeated hundreds to thousands of times. Generally, a counter variable is used to keep track of the number of artificial surveys that lead to a decision that the exposure profile was acceptable. Guidance can be found in Hewett (2005) on how to use this information when designing or fine tuning an exposure assessment strategy.

7.2 Calculating the Fraction of Noncompliant Workers

Let us hypothesize a situation where the L_{TWA} is 1 ppm (as in the OSHA benzene standard), the group geometric standard deviation D is 2, and the 95th percentile for the group exposure profile is exactly equal to the L_{TWA} (i.e., the group exceedance fraction θ is 0.05). Furthermore, let us assume that the group heterogeneity coefficient (ρ) is a moderate 0.20.

This scenario is depicted in Figure 1. The equations for calculating the probability density functions in Figure 1 are in Appendix C. Two curves are shown. One is the distribution of exposures for the entire group, and the other is the *statistical distribution* of "worker 95th percentiles" exposures (see footnote c). Notice that a sizable fraction of the "worker 95th percentile" distribution exceeds the L_{TWA}. This fraction can be estimated using Equation 8. First, we calculate D_w and D_b using Equations 1 and 2: D_w=1.8589 and D_b=1.3634. Second, calculate G using either Equation 12 or Equations 10 and 11. For example, from Equation 10 we determine that a θ of 0.05 corresponds to a Z-value of 1.645. Inserting this value and the L_{TWA} into Equation 11 results in G=0.3197. We can now calculate θ_P :

$$\begin{aligned}\theta_P &= 1-\Phi\left[\frac{\ln 1-\left[\ln 0.3197+1.645\cdot\ln 1.8589\right]}{\ln 1.3634}\right]\\ &= 1-\Phi\left[0.3887\right]\\ &= 1-0.65 = 0.35\end{aligned}$$

In this scenario, given our assumptions, 35% of the workers will have individual 95th percentiles that exceed the L_{TWA} , even though the group 95th percentile equals the limit.

Let us take a different group and assume that the exposure limit in question is a long-term average limit, or L_{LTA} , set at 3 ppm (as in the pre-2002 European vinyl chloride standard; see Footnote a). Let us also assume that the group mean is exactly equal to the L_{LTA} , and that the group geometric standard deviation and heterogeneity coefficient are approximately 2 and 0.20. Estimate the fraction of workers that have individual means exceeding the exposure limit.

This scenario is depicted in Figure 2. Notice that a sizable fraction of the "worker means" distribution exceeds the L_{LTA} of 3 ppm. This fraction can be estimated using Equation 9. As before, we first calculate D_w and D_b using Equations 1 and 2. Second, calculate G. Since we assumed that the group mean is equal to the L_{LTA} , Equation 13 can be used to calculate the group geometric mean, resulting in G=2.3593. We can now calculate θ_M:

$$\begin{aligned}\theta_M &= 1-\Phi\left[\frac{\ln 3-\left(\ln 2.3593+\frac{1}{2}\cdot(\ln 1.8589)^2\right)}{\ln 1.3634}\right]\\ &= 1-\Phi\left[0.1550\right]\\ &= 1-0.56 = 0.44\end{aligned}$$

Given our assumptions, 44% of the workers have individual means that exceed the L_{LTA} , even though the group mean equals the limit.

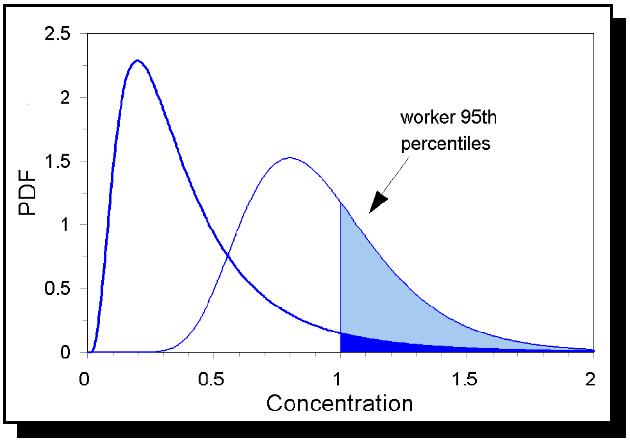

Figure 1: Distribution of worker exposures and distribution of worker "95th percentiles". The geometric mean, geometric standard deviation, and heterogeneity coefficient for the group exposure profile are 0.3197, 2.0, and 0.2, respectively. Exactly 5% of the group exposure profile exceeds the TWA exposure limit of 1 ppm. In contrast, 35% of the workers have individual 95th percentile exposures that exceed the limit.

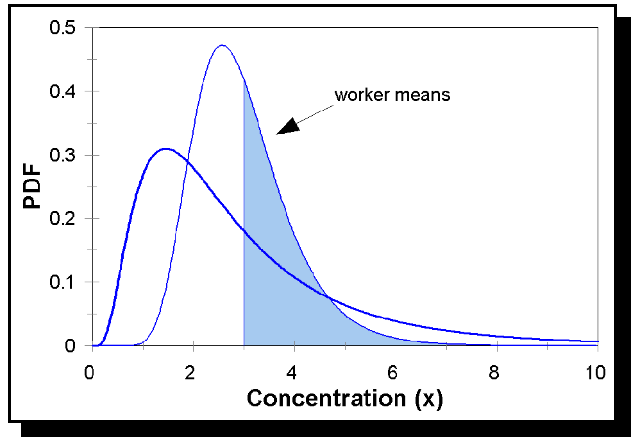

Figure 2: Distribution of worker exposures and distribution of "worker means". The geometric mean, geometric standard deviation, and heterogeneity coefficient for the group exposure profile are 2.3593, 2.0, and 0.2, respectively. The mean of the group exposure profile exactly equals the long-term average limit of 3 ppm. In contrast, 44% of the workers have individual long-term mean exposures that exceed the limit.

7.3 Estimating the Effect of Exposure Reduction

The COV model can be used to estimate the effect on θ_P (or θ_M) of a reduction in either the central tendency or variability of the group exposure profile. For example, say an employer has a continuous improvement program that also applies to health and safety matters. Given the scenario previously discussed for TWA limits (see Section 7.2) approximately 35% of the workers are expected to have 95th percentile exposures that exceed L_{TWA}, even though the group 95th percentile exactly equals the limit. What would be the effect if the employer set a goal to reduce exposures by 10% per year until exposures are acceptable. Let us assume that this will result in a reduction of the group geometric mean G by 30% over the next three years. Assuming that both D and ρ remain roughly the same at 2 and 0.2, θ_P would be expected to decrease from 0.35 to 0.06.

Let us assume the same type of scenario of the long-term exposure limit. Given our assumptions (see Section 7.2), 44% of the workers are expected to have individual means that exceed the L_{LTA} , even though the group mean equals the limit. As before, we will assume that the employer will be able to reduce exposures 10% per year for the next three years. The effect of reducing the group geometric mean by 30%, with D and ρ remaining the same at 2 and 0.2, would be a reduction in θ_M from 0.44 to 0.10.

In both cases, a modest 10% reduction in the central tendency (i.e., group geometric mean exposure) per year over the course of several years should result in a sharp reduction in the fraction of workers that are not in compliance with the exposure imit. The effect of any reductions in either within- and between-worker variability may also be gauged in the same manner.

7.4 AIHA "critical exposure group" Concept

Mulhausen and Damiano (1998) presented the concept of a "critical exposure group"; that is, an exposure group where the sample estimate of the group 95th percentile is greater than half, but less than or equal to the L_{TWA}. They suggested that in this situation the employer should look closely at the definition of the exposure group. They point out that the closer the sample estimate of the group 95th percentile is to the TWA limit, the greater the possibility that an unacceptable fraction of the workers will have individual 95th percentiles in excess of the limit. Calculations such as those above can assist industrial hygienists to conceptualize a "critical exposure group" (see Figures 1 and 2).

7.5 Limitations of the COV Model

The COV model is potentially very useful, but it requires several assumptions:

- First, the model requires that we assume that both the group exposure profile and the individual worker exposure profiles can be described by the lognormal distribution.
- Second, the model requires the assumption that D_w is identical for all workers.
- Last, we assume that there are sufficient workers so that the distribution of individual geometric means is nearly continuous.

All of these assumptions will be violated to some degree, so the COV model will not exactly match the group exposure profile and individual worker exposure profiles for any exposure group:

- The actual exposures are not lognormally distributed, but are only approximately lognormal.
- It is highly improbable, if not impossible, for all workers in any exposure group to have identical within-worker geometric standard deviations.
- The distribution of worker geometric means will approximate a continuous distribution only for very large exposure groups.

However, this model is useful in that it assists us in characterizing and quantifying the components of variability, and in predicting the effectiveness of this or that exposure sampling strategy (Hewett, 1999, 2005).

7.6 Estimating the Model Parameters

It is possible to estimate G, D, and ρ for any exposure group, provided that repeat measurements are available for the majority of the workers (Rappaport, 2000). Using the sample estimates of these parameters one can then calculate the point estimates of other model parameters, such as $X_{0.95}$, $P_{0.95}$, $M_{0.95}$, θ, θ_P, or θ_M. For example, Symanski *et al*. (2001) estimated θ and θ_M for a group of workers.[i] However, procedures for calculating confidence intervals about these estimates have not been developed, except in one instance. Lyles et al. (1997a) reported an *ad hoc* procedure that is basically equivalent to calculating an approximate 90% upper confidence limit on θ_M, but cannot easily be extended to other confidence intervals or applied to the potentially more useful estimates of θ_P.[j] Here there appears to be an opportunity for additional research and development.

7.7 Training

The COV model has an obvious application in the training of industrial hygienists and other occupational health professionals. It can be used to acquire an understanding of the relationships between single shift TWA (and long-term) exposure limits, group exposure profiles, and individual worker exposure profiles.

8 CONCLUSIONS

A conceptual model for describing the exposure profile of an exposure group was presented. The procedure for using this model to generate random exposures was discussed. These random exposures can be used in computer simulations to evaluate the performance of a specific, corporate exposure assessment strategy. Equations were also derived from the model that give insight into the relationship between the group exposure profile and the exposure profiles for individual workers and permit the estimation of true compliance with both single shift and long-term exposure limits.

[i] In Symanski *et al.* (2001) the variables γ and θ are equivalent to θ and θ_M, respectively.

[j] In addition, the Lyles *et al*. (1997a) procedure does not work well whenever there are moderate to low levels of group heterogeneity, forcing the employer to collect large numbers of repeat measurements per worker in order to be able to pass their test (Lyles *et al*., 1977b).

9 ACKNOWLEDGMENTS

I would like to thank Dr. Mark Nicas (University of California) and Dr. Michael Attfield (NIOSH) for their statistical reviews and recommendations. I especially thank Dr. Nicas for the many comments that over the years helped lead to this paper.

---

10 REFERENCES

Aitchison, J. and Brown, J.A.C.: *The Lognormal Distribution with special reference to its uses in economics*. Cambridge University Press (1957).

Hewett, P.: Letter to the Editor - Comments relating to Tornero-Velez *et al*: Compliance Versus Risk in Assessing Occupational Exposures. *Risk Analysis* 18:665-667 (1998a).

Hewett, P.: Letter to the Editor - Comments relating to Lyles, Kupper, and Rappaport: A Lognormal Distribution-based Exposure Assessment Method for Unbalanced Data. *Annals of Occupational Hygiene* 42:413-422 (1998b).

Hewett, P.: Estimating the efficiency and effectiveness of exposure monitoring programs through computer simulation. abstract number 218. presented at the 1999 American Industrial Hygiene Conference and Exposition (Toronto, June 5-11) (see also www.aiha.org/abs99/9expo.html) (1999).

Hewett, P.: Misinterpretation and Misuse of Exposure Limits. *Applied Occupational and Environmental Hygiene* 16:251-256 (2001).

Hewett, P.: Performance-based exposure assessment strategies for TWA exposure limits. Technical Report No. 05-03. Exposure Assessment Solutions, Inc. (www.oesh.com) (2005).

HSC (Health and Safety Commission): Proposals for amending the Control of Substances Hazardous to Health Regulations 2002. United Kingdom Health and Safety Executive (2002).

Kromhout, H., Symanski, E., and Rappaport, S.M.: A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. *Annals of Occupational Hygiene* 37:253-270 (1993).

Leidel, N.A., Busch, K.A., Lynch, J.R.: *Occupational Exposure Sampling Strategy Manual*. National Institute for Occupational Safety and Health (NIOSH) Publication No. 77-173 (available from the National Technical Information Service (NTIS), Publication No. PB274792) or as the NIOSH website as downloadable .pdf file) (1977).

Leidel, N.A. and Busch, K.A.: Chapter 52 - Statistical Design and Data Analysis. *Patty's Industrial Hygiene, Fifth Edition, Volume 4.* Edited by R.L. Harris. John Wiley & Sons, Inc. (2000).

Lyles, R.H., Kupper, L.L., and Rappaport, S.M.: A Lognormal Distribution-based Exposure Assessment Method for Unbalanced Data. *Annals of Occupational Hygiene* 41:63-76 (1997a).

Lyles, R.H., Kupper, L.L., and Rappaport, S.M.: Assessing Regulatory Compliance of Occupational Exposures Via the Balanced One-Way Random Effects ANOVA Model. *Journal of Agricultural, Biological, and Environmental Statistics* 2:64-86 (1997b).

Mulhausen, J. and Damiano, J.: (editors): *A Strategy for Assessing and Managing Occupational Exposures, Second Edition*. Fairfax, VA: American Industrial Hygiene Association (1998).

Nicas, M. and Spear, R.C.: A Probability Model for Assessing Exposure Among Respirator Wearers. Part II- Overexposure to Chronic Versus Acute Toxicants. *American Industrial Hygiene Association Journal* 53:419-426 (1992).

Oldham, P.D., and Roach, S.A.: A Sampling Procedure For Measuring Industrial Dust Exposure. *British Journal of*

Industrial Medicine :9:112-119 (1952).

Petersen, M.R., Sanderson, W.T., and Lenhart, S.W.: Application of a Pilot Study to the Development of an Industrial Hygiene Sampling Strategy. *American Industrial Hygiene Association Journal* 47:655-658 (1986).

Rappaport, S.M.: Interpreting Levels of Exposures to Chemical Agents. in *Patty's Industrial Hygiene and Toxicology, Fifth Edition, Volume 1*.; Harris, R.L., editor; John Wiley & Sons, Inc. (2000).

Rappaport, S.M., Lyles, R.H., and Kupper, L.L.: An exposure-assessment strategy accounting for within- and between-worker sources of variability. *Annals of Occupational Hygiene* 39:469-495 (1995).

Spear, R.C., Selvin, S., and Francis, M.: The Influence of Averaging Time on the Distribution of Exposures. *American Industrial Hygiene Association Journal* 47:365-368 (1986).

Spear, R.C., Selvin, S., Schulman, J., and Francis, M.: Benzene exposure in the petroleum refining industry. *Applied Industrial Hygiene* 2:155-163 (1987).

Spear, R.C. and Selvin, S.: OSHA's Permissible Exposure Limits: regulatory compliance versus health risk. *Risk Analysis* 9:579-586 (1989).

Spear, R.C.: Individual versus group differences in exposure and risk. in *Exposure Assessment for Epidemiology and Hazard Control*, S.M. Rappaport and T.J. Smith, editors; American Conference of Governmental Industrial Hygienists Industrial Hygiene Science Series, Lewis Publishers (1991).

Symanski, E., *et al*.: Heterogeneity in Sources of Exposure Variability Among Groups of Workers Exposed to Inorganic Mercury. *Annals of Occupational Hygiene* 45:677-687 (2001).

Tornero-Velez, R., Symanski, E., Kromhout, H, Yu, R.C., and Rappaport, S.M.: Compliance Versus Risk in Assessing Occupational Exposures. *Risk Analysis* 17:279-292(1997).

Tuggle R.M.: The relationship between TLV-TWA compliance and TLV-STEL compliance. *Applied Occupational and Environmental Hygiene* 15:380-386 (2000).

11 APPENDIX A - COMPONENTS OF VARIANCE MODEL

Assume that the exposure (x) is lognormally distributed and let *y=ln(x)*. We know that for any exposure group the variability in exposures can be divided into within-worker and between-worker components. Let us assume that the *jth* exposure for the *kth* worker is a function of the group mean (of the log-transformed exposures), a between-worker component, and a within-worker component:

$$y_{jk} = \mu_y + \beta_k + \varepsilon_{jk}$$

where μ_y is the true group mean (of the log-transformed exposures), β_k is a worker effect and reflects the deviation from the overall mean for the *kth* worker, and ε_{jk} is the within worker variation. We further assume that $\beta_k \sim \mathbb{N}(0, \sigma_b^2)$, and $\varepsilon_{jk} \sim \mathbb{N}(0, \sigma_w^2)$.\k

The exposure profile for the *kth* worker is

$$y_{jk} \sim \mathbb{N}(\mu_y + \beta_k, \sigma_w^2)$$

and the exposure profile for the group is

$$y_{jk} \sim \mathbb{N}(\mu_y, \sigma_b^2 + \sigma_w^2).$$

For the exposure group, the expected value of the *jth* (untransformed) exposure for the *kth* worker is the true group mean:

$$E[x_{jk}] = \mu = \exp\left[(\mu_y + (\sigma_b^2 + \sigma_w^2)/2\right].$$

For a given worker, the expected value of the *jth* untransformed exposures is the true worker mean:

$$E[x_{jk}] = M_k = \exp\left[(\mu_y + \beta_k) + \sigma_w^2/2\right]$$

Since μ_y and σ_w^2 are constants, and $\beta_k \sim \mathbb{N}(0, \sigma_b^2)$, the log-transformed worker means have the following distribution:

$$\ln(M_k) \sim \mathbb{N}(\mu_y + \sigma_w^2/2, \sigma_b^2)$$

Let P_k represent the 95th percentile on the untransformed scale for the *kth* worker. Since $\ln(P_k) = (\mu_y + \beta_k) + 1.645 \cdot \sigma_w$, it follows that the log-transformed worker 95th percentiles have the following distribution:

$$\ln(P_k) \sim \mathbb{N}(\mu_y + 1.645 \cdot \sigma_w, \sigma_b^2)$$

\k $\mu_y = \ln(G)$, $\sigma_b = \ln(D_b)$, and $\sigma_w = \ln(D_w)$.

12 APPENDIX B - GROUP HETEROGENEITY COEFFICIENT

Kromhout *et al*. (1993) studied the distribution of ρ amongst 165 exposure groups. The median value of ρ was 0.20, with a lower quartile value of 0.05 and an upper quartile value of 0.40. We will provisionally define the following ranges of ρ for indicating low, medium, and high heterogeneity exposure groups:

Low - $\rho \leq 0.10$
Med - $0.10 < \rho \leq 0.30$
High - $\rho > 0.30$

For convenience, we will use the following values of ρ to indicate low, medium, and high between-worker variability:

Low - $\rho = 0.05$
Med - $\rho = 0.20$
High - $\rho = 0.40$

The study of Kromhout *et al*. (1993) gives us insight into the range of values to be expected and permits provisional definitions of low, medium, and high group heterogeneity. But for a specific process or operation that is common throughout a particular industry, it would be useful to know the range of group heterogeneity coefficients ρ for different types of controls or work practices. Given this information, and estimates of a group geometric mean G and geometric standard deviation D, one could use computer simulation to estimate the degree of actual compliance with either a single shift or long-term exposure limit.

13 APPENDIX C - CALCULATION OF PROBABILITY DENSITY FUNCTIONS

The following equations can be used to calculate and plot the probability density functions seen in Figures 1 and 2. Each can easily be calculated using the spreadsheet function for the normal distribution. For example, the group probability density function (pdf) can be calculated as "1/x*normdist(ln(x),ln(G),ln(D),0)", where x is greater than zero. The other functions are calculated in a similar manner.

Group probability density function (i.e., group exposure profile):

$$pdf(x) = \frac{1}{x \cdot \ln D \sqrt{2\pi}} \exp\left[\frac{-(\ln x - \ln G)^2}{2(\ln D)^2}\right]$$

Probability density function for the "worker geometric mean exposures":

$$pdf(G_k) = \frac{1}{G_k \cdot \ln D_b \sqrt{2\pi}} \exp\left[\frac{-(\ln G_k - \ln G)^2}{2(\ln D_b)^2}\right]$$

Probability density function for the "workers 95th percentile exposures":

$$pdf(P_k) = \frac{1}{P_k \cdot \ln D_b \sqrt{2\pi}} \exp\left[\frac{-(\ln P_k - \ln G_P)^2}{2(\ln D_b)^2}\right] \qquad \text{where } \ln G_P = \ln G + 1.645 \cdot \ln D_w.$$

Probability density function for the "worker mean exposures":

$$pdf(M_k) = \frac{1}{M_k \cdot \ln D_b \sqrt{2\pi}} \exp\left[\frac{-(\ln M_k - \ln G_M)^2}{2(\ln D_b)^2}\right] \qquad \text{where } \ln G_M = \ln G + 1/2 \cdot (\ln D_w)^2.$$