SCI 5119-1

SEQUENTIAL SAMPLING PLANS AND DECISION THEORY FOR EMPLOYER MONITORING OF EMPLOYEE EXPOSURE TO INDUSTRIAL ATMOSPHERES

FINAL REPORT

CONTRACT NO. CDC-99-74-75

MAY 1975

SCI SYSTEMS CONTROL, INC.

SYSTEMS CONTROL, INC.
1801 PAGE MILL ROAD
PALO ALTO, CALIFORNIA 94304

TELEX: 348433

TELEPHONE (415)
494-1165

May, 1975

Final Report
SCI Project 5119

SEQUENTIAL SAMPLING PLANS AND DECISION THEORY FOR EMPLOYER MONITORING OF EMPLOYEE EXPOSURE TO INDUSTRIAL ATMOSPHERES

CONTRACT CDC-99-74-75

Prepared for:

National Institute for Occupational
Safety and Health
Post Office Building
Cincinnati, Ohio 45202

Prepared by:

Y. Bar-Shalom
D. Budenaers

Approved by:

Dr. R. B. Schainker
Manager, Environmental Systems Program

ACKNOWLEDGEMENT

The authors acknowledge the encouragement of the contract officers Messrs. Nelson Leidel and Kenneth Busch, as well as Mr. Jeremiah Lynch, from the National Institute for Occupational Safety and Health, Cincinnati, Ohio. Their suggestions and stimulating discussions conducted during this contract were helpful and constructive. Thanks are also due to Mr. James W. Patmore of Systems Control, Inc., for his helpful comments and suggestions.

ABSTRACT

This report presents the results of a study conducted at Systems Control, Inc., Palo Alto, Ca., and sponsored by the National Institute for Occupational Health and Safety, Cincinnati, Ohio under Contract CDC-99-74-75. The purpose of the study was to develop a procedure to be used by employers to ensure the employees' safety in an industrial environment where atmospheric contaminants are present. A sequential procedure is presented according to which it is recommended that the employer sample the employees potentially exposed to atmospheric contaminants. Statistical decision procedures for classifying the employees according to their exposure are presented. Also the possible action to be taken by the employer, namely instituting engineering controls if warranted, is discussed and the corresponding decision procedure is derived.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND NOTATIONS

Section 3

n - number of samples from a group

N - group size

τ - top fractile

N_o - top subgroup size

p_o - probability of missing a member from a subgroup

α - allowed probability of miss

Subsection 4.1

α - probability of error of type I

α' - probability of error of type II

LCL - lower confidence limit

UCL - upper confidence limit

Subsection 4.4

i - index of measurements within a day

X_j - ceiling measurements on a given day

CSTD - ceiling standard

x_j - ceiling measurements normalized (divided) by the CSTD

n - number of ceiling measurements from a given day

y_i - common logarithm of x_i

y^n - set of available data $\{y_i, \ldots, y_n\}$

$\mathcal{N}$ - normal (Gaussian) probability density function

μ - population mean of y_i

σ - population standard deviation of y_i

$\bar{y}$ - sample mean of y_i

p_c - probability of compliance

β - probability that one ceiling value exceeds the CSTD

s - sample standard deviation of y_i

n' - number of critical intervals

$x \sim \mathcal{N}(\mu, \sigma^2)$ - The random variable x is normally distributed with mean μ and variance σ^2.

Subsection 4.5

i - index of days

x_i - daily composite measurement on day i

μ_i - true value of one 8-hour average on day i

σ_i - standard deviation of the error in x_i

$\overline{x}_n$ - sample average of x_i

$\hat{x}_n$ - modified average of x_i

$\hat{\mu}_n$ - mean of $\hat{x}_n$

$\hat{\sigma}_n$ - standard deviation of $\hat{x}_n$

γ - gain of recursive estimator

β - discount rate of estimator

$\hat{\lambda}_n$ - discounted least squares estimate of the multiday average

VAR - variance

Subsection 5.1

p_v - probability of violation

$\tilde{p}$ - probability threshold

Subsection 5.2

w_i - measurement noise

p_i - observed probability of violation on day i

π_v - common log of p_v

π_i - common log of p_i

$\hat{\pi}_n$ - estimate of π_v

$\hat{\tau}_n$ - standard deviation of $\hat{\pi}_v$

$\tilde{\pi}$ - common log of $\tilde{p}$

Subsection 5.3

- $\bar{y}$ – sample mean of common logs of the normalized measurements (grab samples)
- s – sample standard deviation of the same measurements
- a – natural logarithm of 10
- $\bar{y}_T$ – true value of the means of the common logs of the normalized measurements
- σ_T – true value of the standard deviation of the above measurements
- m_i – true 8-hour exposure
- n_i – number of available grab samples
- E – expectation operator
- w.r.t. – with respect to

Section 6

- x_i – daily composite measurement on day i
- GSD – geometric standard deviation
- p_v – probability of violation
- σ_i – standard deviation of x_i

1. INTRODUCTION

This report presents the results of a study aimed at developing a procedure to be used by employers to ensure their employees' safety in an industrial environment where atmospheric contaminants are present. The main objective was to develop a procedure that can be implemented with a minimum of burden on the employer. The detailed step-by-step procedure is presented in a companion Handbook [B2]. The focus of this report is on the underlying assumptions and mathematical derivations of the corresponding statistical decision procedures.

1.1 SUMMARY OF STUDIES DONE UNDER THIS CONTRACT

The overall procedure the employer has to carry out has been formulated in a sequential manner and the subprocedures it consists of have been identified. In line with the requirement of minimum burden on the employer, the population of workers is proposed to be divided into groups with similar exposure and a partial sampling procedure is to be carried out within each group. The partial sampling procedure has been developed to satisfy the following requirement: at least one worker from a given top fractile in terms of exposure has to be included in the sample with a given high probability. Once the sampling results are available, statistical decision theory methods are used to classify the employees as overexposed, exposed and unexposed. The mathematical formulation of these groups in terms of confidences is presented, as well as the procedure to carry out the corresponding hypothesis testing problem. A recursive estimation method for the multi-day exposure has been developed and the associated confidence region is also discussed. This estimate of the multi-day exposure can serve as an indication of the potential effect of substances that have a cumulative effect. A procedure for deciding whether engineering controls are to be instituted has been devised. The criterion for instituting controls is the following: They are to be instituted if there is a high confidence that the maximum allowable contaminant concentration as set forth by the Federal Standard [F1] is exceeded, in the long run, more than a given fraction of the time. This decision is made based upon the "probability of violation" (violation is an excess of the standard). This probability, which is the best indication of the fraction of time the standard is exceeded, is estimated from the available samples. Finally, a graphical procedure is presented to decide when new samples are to be taken, based upon how close to the standard the latest samples have been.

1.2 ORGANIZATION OF THIS REPORT

Section 2 presents the overall procedure the employer has to carry out to ensure the safety of the employees in an industrial environment with atmospheric contaminants. The various subprocedures are briefly described in this section; their detailed presentation follows in the remaining sections. The partial sampling procedure is the topic of Section 3. This section contains tables of the required sample sizes for several combinations of top fractiles and desired confidences. Section 4 deals with the decision on the exposures. First the classification into the three categories (overexposed, exposed, and underexposed) is given in Subsection 4.1. The decisions on the 8-hour exposure are only briefly discussed in Subsections 4.2 and 4.3 because they are documented in earlier studies. A detailed presentation of the decision procedure on the ceiling exposure developed in the course of the present study is given in Subsection 4.4. This section is concluded with a discussion of the multi-day average exposure in Subsection 4.5. The topic of Section 5 is the decision on instituting engineering controls. The problem is formulated mathematically in Subsection 5.1 in terms of the probability of violation. The procedure to obtain the estimate of the probability of violation and the associated confidence region is discussed for the long-term samples case in Subsection 5.2. The grab samples case is presented in Subsection 5.3. The recommended intervals for sampling are the topic of Section 6.

2. THE SAMPLING STRATEGY FOR THE EMPLOYER

A sequential procedure to be used by an employer to ascertain the safety of the employees, called sampling strategy, has been designed with the following objectives:

a. To provide statistical methods for deciding whether the employees are safe,

b. To simplify the implementation in order to make its use widespread,

c. To minimize the sampling burden on the employer while complying with the law.

The flowchart in Fig. 2.1 describes the steps to be taken by the employer to ensure employee safety while minimizing the sampling burden. A brief description of each block in the flowchart is given next.

1. The initial determination of groups of workers with similar exposure is performed using industrial hygiene considerations. These are presented in detail in the Handbook [B2], Section 4.1. The purpose of dividing the employees into groups of similar exposure is to enable the sampling to proceed sequentially, starting from the group with the highest exposure. This first group will consist of the "employee with the highest risk" or, if no single employee can be initially determined as such before measurements are taken, then it will consist of N_1 "highest risk" employees.

2. If the group consists of a large number of employees with similar exposure, then, in order to minimize the sampling burden, a subgroup will be first sampled. The criterion is that with a high confidence, at least one worker from the, say, top 10% (in terms of exposure) will be sampled. The theory behind this partial sampling procedure is presented in Section 3.

3. After the samples have been obtained and analyzed and the results recorded, the workers are to be classified into the following categories: unexposed (Category C, if there is a high confidence that the exposure is below the Federal Standard), overexposed (Category A, if there is a high confidence that the exposure exceeds the standard), and exposed (Category B, if neither of the above holds). These three classifications are discussed in detail in Section 4. The classifications (decisions) are made for both 8-hr. average as well as ceiling values, as applicable. The decision procedure on the 8-hr. average exposure based upon long-term (>1 hr.) measurements was presented in Leidel and Busch [L1]; for the case where short-term, i.e., grab samples (>1/2 hr.) only are available the procedure was derived in Bar-Shalom, et al. [B1]. The decision concerning ceiling levels is presented in Section 4. All of these decision procedures are documented in a user-oriented fashion in the Handbook [B2].

4. If, based upon the obtained samples, the employee with the highest observed exposure is classified as unexposed, then one proceeds to block 5 and the sampling of the current and lower groups is not necessary.

5. No more samples are to be taken from the group under consideration and the lower groups for a period of six months unless changes in the plant take place. The workers classified as overexposed (Category A) or exposed (Category B), if any, are to be sampled according to the recommended Sampling Policy (block 8; this is discussed in detail in Section 6).

6. Unless all the employees from the subgroup being sampled are unexposed, the whole group has to be sampled.

7. The decision on the institution of engineering controls is to be made using the following criterion: Institute controls if there is a high confidence that the workers are exposed above the

standard more than a given percentage of the time. The procedure to obtain this decision is presented in Section 5.

8. The recommended Sampling Policy, consisting of guidelines for the determination of the period after which new samples are to be taken, is presented in Section 6. The calculation of the recommended interval between samples depends on how near the standard the latest measurements have been.

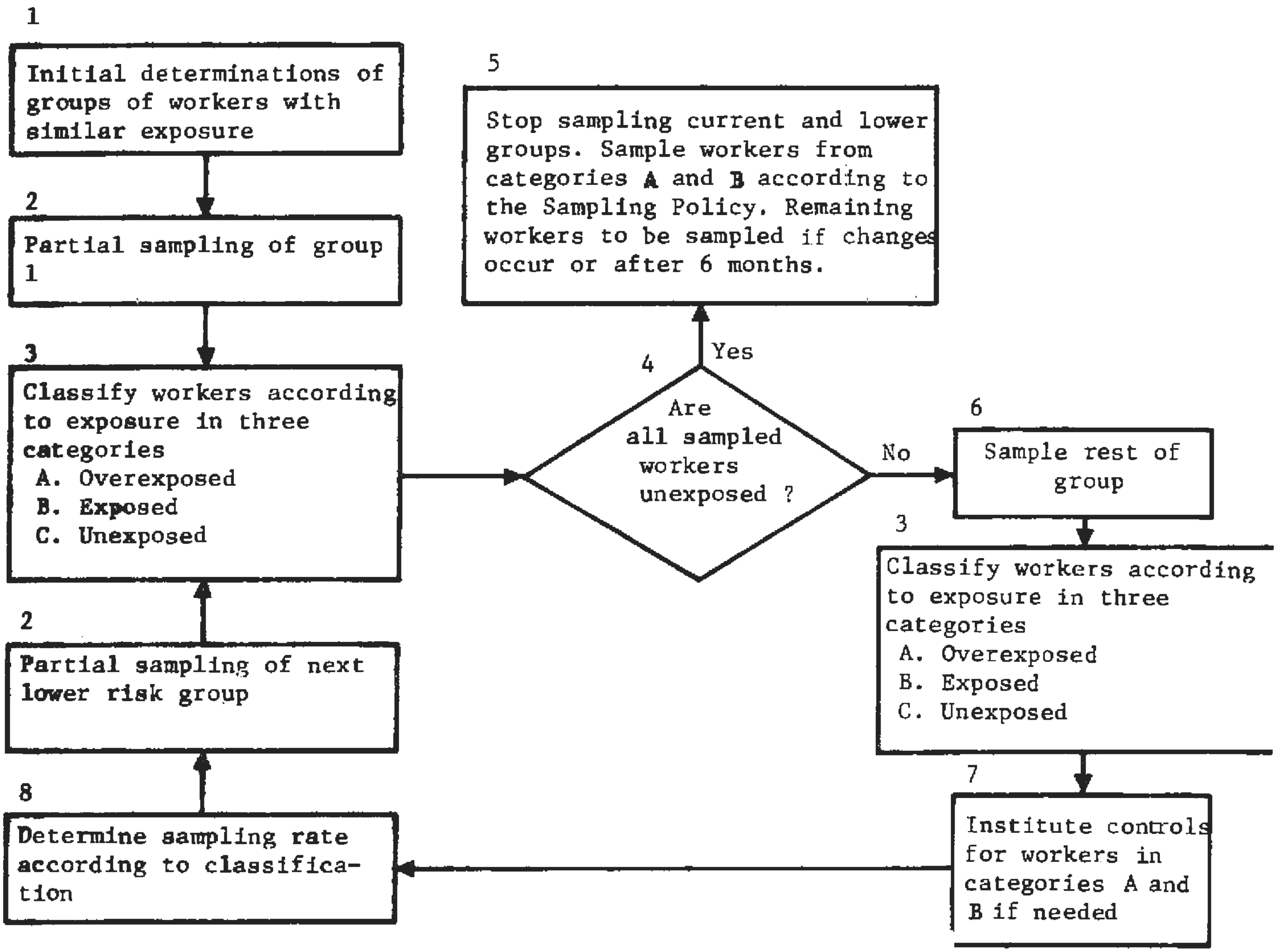

Figure 2.1 Sampling Strategy for Employer

3. THE PARTIAL SAMPLING PROCEDURE

This section describes a partial sampling procedure that can be used by an employer in order to minimize the sampling burden. The procedure is applicable in the case where there is a sizable group of workers who, based upon prior information, are similarly exposed. The number of workers in such a group is denoted by N and a random sample of $n < N$ is to be taken.

The criterion will be that at least one worker from those with high exposure should be in this sample with high probability. If "high exposure" is defined as the "top 10%" then the sample will have to include (with high probability, say, $1 - \alpha$) one worker out of a given subgroup of size, say, $N_o = \tau N$ where τ is the top fractile, $0 < \tau < 1$. In the "top 10%" case, $\tau = 0.1$. The allowed probability of missing all workers from the top N_o in the sample of n out of N is α.

The expression of the probability of missing all workers from a subgroup of size N_o from a group of N when sampling n is

$$p_o = \frac{(N - N_o)!}{(N - N_o - n)!} \quad \frac{(N - n)!}{N!} \qquad (3.1)$$

This expression follows from calculations in the theory of sampling without replacement treated, e.g., in [P1]. Note that

$$p_o = p_o(N,\tau,n) \qquad (3.2)$$

and in order to obtain the sample size the following equation has to be solved

$$p_o(N,\tau,n) = \alpha \qquad (3.3)$$

for the sample size n, given N (the size of the group under consideration), τ (the desired top fractile), and α (the allowed probability of miss).

The solution, rounded off to the nearest integer is presented in Tables 3.1 - 3.4 for the following ranges of values:

a. Group size $N = 1, \ldots, 50$

b. Top 10% and 20%, i.e., $\tau = .1, .2$

c. Confidence levels of 90% and 95%, i.e., $\alpha = .1, .05$.

The solution for very large N is obtained from the sampling with replacement problem because in this case $n \ll N$. The procedure in this case is to guarantee with confidence $1 - \alpha$ that in n trials at least one event, whose probability of occurring in one trial is τ, will occur. Thus the probability of such an event not occurring in n trials has to be α, i.e.,

$$(1 - \tau)^n = \alpha \tag{3.4}$$

and

$$n = \frac{\log \alpha}{\log (1-\tau)} \tag{3.5}$$

For example,

$$n(\tau = .1, \quad \alpha = .1) \approx 22 \tag{3.6}$$

and this is the limit towards which n tends in Table 3.1 as $N \to \infty$.

Note that even for $N = 50$ the value of n from Table 3.1 is still far from the above limit and thus it is important to use the sampling without replacement approach, i.e., (3.3).

Size of group N	8	9	10	11-12	13-14	15-17	18-20	21-24	25-29	30-37	38-49	50
Number of required samples n	7	8	9	10	11	12	13	14	15	16	17	18

Table 3.1. Partial sampling procedure for top 10% and confidence .90 (n = N if N $\leq$ 7)

Size of group N	12	13-14	15-16	17-18	19-21	22-24	25-27	28-31	32-35	36-41	42-50
Number of required samples n	11	12	13	14	15	16	17	18	19	20	21

Table 3.2. Partial sampling procedure for top 10% and confidence .95 (n = N if N $\leq$ 11)

Size of group N	6	7-9	10-14	15-26	27-50
Number of required samples n	5	6	7	8	9

Table 3.3. Partial sampling procedure for top 20% and confidence .90 ($n = N$ if $N \leq 5$)

Size of Group N	7-8	9-11	12-14	15-18	19-26	27-43	44-50
Number of required samples n	6	7	8	9	10	11	12

Table 3.4. Partial sampling procedure for top 20% and confidence .95 ($n = N$ if $N \le 6$)

4. DECISION ON EXPOSURES

In this section the decision procedure that results in the classification of the sampled workers based upon their measured exposure is discussed. The definitions of the three classes are presented in Subsection 4.1. The decision on the 8-hr. average exposure is done in two different ways depending upon the type of measurements. Since these two procedures have been developed elsewhere [L1,B1], they are only briefly discussed in Subsections 4.2 and 4.3. The decision regarding the ceiling exposure is presented in Subsection 4.4. The concept of multiday average exposure and the procedure to obtain it are the topics of Subsection 4.5.

4.1 CLASSIFICATION OF EMPLOYEES BASED ON MEASURED EXPOSURE

The workers are to be classified into three categories based upon the measured exposures and the associated confidence.

This classification is equivalent to the 3-way decision between the hypotheses

H_o: exposure below standard

H_1: exposure above standard

subject to maximum allowed probabilities of error of type I and II, denoted as α and α', respectively. The 3-way decision comes from the fact that rejecting one hypothesis does not yet imply accepting the other. This is due to the constraints on both probabilities of error. In other words, only if the probability of type II error is below α', i.e.,

$$P\{"H_o"|H_1\} \leq \alpha' \qquad (4.1.1)^*$$

then the decision is "H_o" and the confidence on the correctness of this decision is at least $1-\alpha'$.

---

* This is read as follows: probability of accepting (declaring) H_o given that H_1 is correct in less or equal to α'.

The decision "H_o" means that with (high) confidence of at least $1 - \alpha'$ the worker's exposure is below the standard.* The corresponding terminology used in [L1, B1] for the null hypothesis H_o was either "compliance" or "no action". In this case the worker is said to be unexposed and this category is denoted as C, the lowest in terms of ranking according to exposure.

Hypothesis H_1 is accepted only if the probability of type I error is below α, i.e.,

$$P\{"H_1"|H_o\} \leq \alpha \qquad (4.1.2)$$

Then the decision "H_1" means that with (high) confidence of at least $1 - \alpha$ the worker's exposure exceeds the standard. The corresponding terminology used in the previous studies [L1,B1] was "violation" or "no compliance". In such a case the worker is said to be overexposed and this category, the highest, is denoted as A.

If no decision can be made subject to the two maximum allowed probabilities of error then the worker is said to be exposed. This corresponds to the "no decision" region in [L1, B1] and this category is denoted as B.

From (4.1.1), a worker belongs to the unexposed category if the upper confidence limit (UCL) at level α' for the exposure is below the standard, i.e.,

$$\text{UCL} < \text{STD} \rightarrow \text{unexposed ("}H_o\text{")}$$

(In the above, the arrow means "implies".)

Similarly, in view of (4.1.2), a worker belongs to the overexposed category if the lower confidence limit (LCL) at level α for his exposure exceeds the standard, i.e.,

$$\text{LCL} > \text{STD} \rightarrow \text{overexposed ("}H_1\text{")}$$

---

*Such procedures are in common use in the area of quality control; see, e.g., Duncan [D1].

These decisions (classifications) are made with one-sided confidence regions illustrated in Fig. 4.1.1. The bullets denote the (daily composite) sample value and the bars show the confidence limit. To stress that the confidence regions are one-sided, an arrow is pointed upwards when we have a LCL and downwards when we have a UCL.

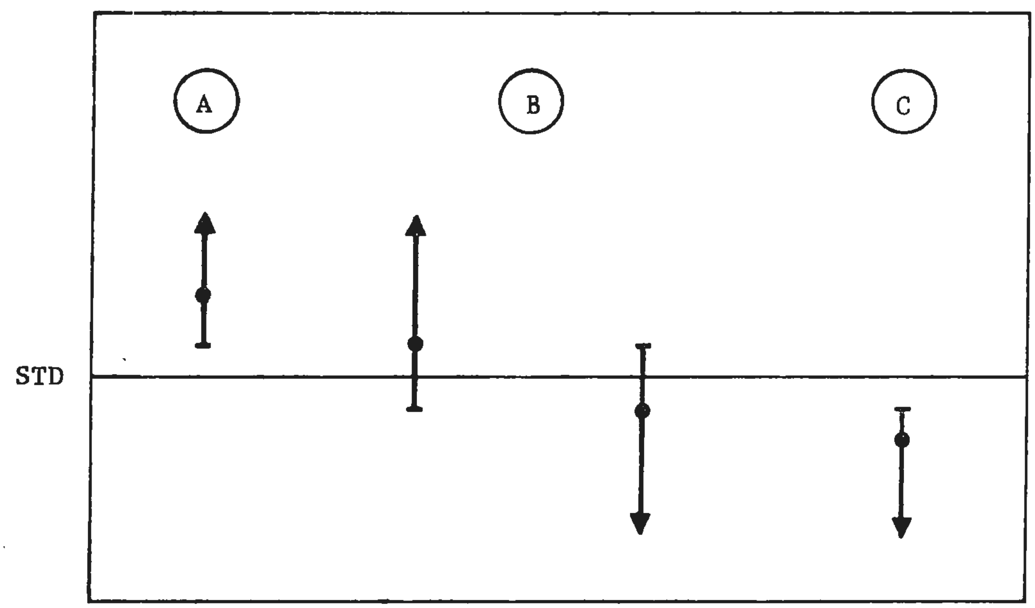

Figure 4.1 Classification According to One-Sided Confidence Regions

4.2 DECISION ON THE 8-HR. EXPOSURE WITH LONG-TERM MEASUREMENTS

If long-term measurements have been taken, the model used is based upon the normality assumption with known accuracy of the measurements. The accuracy corresponds to the sampling/analytical procedure used to obtain the samples. This procedure has been presented in detail in Leidel and Busch [L1] and the corresponding operational steps are described in the Handbook [B2].

4.3 DECISION ON THE 8-HR. EXPOSURE BASED UPON GRAB SAMPLES

If grab samples have been taken, the model is based upon the lognormal assumption with both mean and variance unknown. The decision procedure for this case has been developed in an earlier study [B1] and is to be carried out using decision charts that test the mean of the lognormal distribution. The detailed operational steps necessary to carry out this decision are presented in the Handbook [B2] together with the appropriate decision charts.

4.4 DECISION ON CEILING EXPOSURES

The problem in the ceiling decision procedure is the following: given a set of samples of (usually 15 minute) ceiling level measurements on a day, an inference has to be made about

a) The exposure during the observed intervals,

b) The exposure during the remaining (unobserved) intervals of that day.

a. Decision on the Exposure During the Observed Intervals

The decision about the exposure for the sampled intervals is performed by using the one-sided confidence region for the highest observed value. This confidence region is obtained assuming the measurement error as normally distributed with known standard deviation. This standard deviation is available from the coefficient of variation of the sampling/analytical procedure. If all the available samples indicate that the exposure during the observed intervals is (with high confidence) below the ceiling standard (CSTD) then one has to proceed to (b) and make a statistical inference for the remaining unsampled intervals ("potential samples"). This procedure is described next.

b. <u>Decision on the Exposure During the Remaining Intervals</u>

A test of the following hypothesis

H_o: The whole population of potential samples is below the ceiling standard (CSTD).

vs. the alternative

H_1: At least one of the potential samples could exceed the CSTD

is to be performed.

Assume the following set of ceiling measurements (i.e., of duration equal to the one for which the ceiling standard has been set) from a given day is available: X_j, j = 1, ..., n. Let

$$x_j = \frac{X_j}{CSTD} \tag{4.4.1}$$

be the normalized (with respect to the ceiling standard) measurements.

Since the samples are short-term ones, and if they are noncontiguous, then it can be assumed that they are i.i.d. (independent identically distributed) lognormal random variables [L1, B1]. Furthermore, since in this case, temporal variations only are being considered, the measurement noise will be neglected.

The statistical model will be formulated in terms of the logarithms (base 10) of the normalized data. Therefore, let

$$y_j = \log x_j, \; j=1, \ldots, n \tag{4.4.1}$$

To make a decision concerning an employee's ceiling level exposure, the following hypotheses must be tested with given maximum probabilities of error of type I and II.

$$H_0 \; : \; y_j \leq 0 \text{ for all } i=n+1,\ldots,N \tag{4.4.2}$$

vs.

$$H_1 \; : \; y_j > 0 \text{ for at least one } i, n+1 \leq i \leq N \tag{4.4.3}$$

where N is the size of the sample space (e.g., if the ceiling level standard is set for 15-minute intervals then N=32 for an 8-hour day). "H_o" is the compliance decision ("unexposed" classification) and "H_1" is the violation decision ("overexposed" classification). If neither decision can be asserted with sufficiently high confidence, then a "no decision" choice is made, i.e., "exposed" classification.

The above hypothesis testing problem can be formulated in terms of a probability statement as follows:

Given the set of samples $y^n \triangleq \{y_1, \ldots, y_n\}$, compute the probability of compliance

$$p_c \triangleq P\{y_{n+1} \leq 0, \ldots, y_N \leq 0 | y^n\} \qquad (4.4.4)$$

The probability density of one of the potential samples can be written as

$$p(y_k | y^n) = \iint p\ (y_k, \mu, \sigma | y^n) d\mu d\sigma, k = n+1 \ldots, N \qquad (4.4.5)$$

where μ and σ are the (unknown) mean and standard deviation of y_j, j=1, ..., N, and $p(y_k, \mu, \sigma | y^n)$ is the joint a posteriori density of y_k, μ and σ given the observations y^n.

Using the fiducial distribution of μ [K1],

$$\mu \sim \mathcal{N}(\bar{y}, \frac{\sigma^2}{n}) \qquad (4.4.6)$$

where $\mathcal{N}(a,b)$ is the normal density with mean a and variance b and

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \qquad (4.4.7)$$

and assuming for time being σ as known, one obtains from (4.4.5)

$$p(y_k|y^n) = \mathcal{N}[\bar{y},\sigma^2(1+\frac{1}{n})] \tag{4.4.8}$$

Then

$$P\{y_k>0|y^n\} = \int_0^\infty \mathcal{N}[y_k;\ \bar{y},\sigma^2(1+\frac{1}{n})]\ dy_k \triangleq \beta$$

$$k=n+1,\ldots,N \tag{4.4.9}$$

The probability of compliance (4.4.4) is now given by

$$p_c = \prod_{k=n+1}^{N} P\{\ y_k\leq 0\}$$

$$= \prod_{k=n+1}^{N} [1-P\{y_k>0\}] \tag{4.4.10}$$

Using the notation introduced in (4.4.9) one has

$$p_c = (1-\beta)^{N-n} \tag{4.4.11}$$

If $(N-n)\beta<< 1$ then a good approximation for the above is

$$p_c = 1-(N-n)\ \beta \tag{4.4.12}$$

The assumption of known σ is not totally justified. An approach that would account for this additional uncertainty could be developed along the lines of [B1] using Bayesian arguments with diffuse priors. However, the resulting procedure is of a complexity that would prevent it from being implemented. In view of this, the sample variance

$$s^2 = \frac{1}{n-1} \sum_{j=1}^{n} (y_j - \bar{y})^2 \qquad (4.4.13)$$

is recommended to be used in (4.4.9) in place of σ.

An inspection of Eq. (4.4.11) leads to the observation that, if N-n, the number of unobserved intervals, is large, the probability of compliance p_c becomes small. This is due to the fact that there are more "chances" for at least one sample to exceed the standard and this is bound to happen if we wait long enough. Therefore, the direct application of (4.4.11) might be overly pessimistic. This leads to the concept of expected number of peaks during a day to be discussed next.

Suppose that a "biased" ceiling sampling procedure was utilized to obtain (at random) a set of samples from the "critical" intervals. From knowledge of the industrial process the number of remaining peaks during the day is available and equal to, say, n'. Then the number of unsampled intervals in Eq. (4.4.9) is taken as n' rather than N-n. If all the peak intervals were sampled then there would be no need to go to the inference procedure for the unsampled intervals and the only test to be done is the one described under (a) at the beginning of this subsection. Recall however, that the motivation for developing the inference procedures based upon samples from only a part of the working day stems from the basic objective of minimizing the employer's burden. Thus, if the available samples have been taken from known peaks and there are an additional n' unsampled peaks during the day, the procedure is as follows. If the available samples do not indicate overexposure or exposure, then the decision (classification) is to be made based upon

$$p_c = (1-\beta)^{n'} \qquad (4.4.14)$$

If the probability of compliance p_c exceeds a preset threshold, say .9, the worker is classified as unexposed. On the other hand, if p_c is below another threshold, say .1, then the worker can be classified as overexposed; otherwise the classification is exposed. For details of the procedure see the Handbook [B2].

4.5 THE MULTI-DAY AVERAGE EXPOSURE

This subsection presents a procedure that combines the contaminant measurements from past days in order to obtain an indication of the exposure over several days. Such an estimate of the exposure over several days should have the following properties:

a. Adaptivity to new data,

b. Smooth response to new data,

c. Insensitive to underlying distribution assumption,

d. Modest (i.e., as little as possible) computational requirements.

First a heuristic development of a recursive method for contaminant estimation is given. This development starts by using the arithmetic average of the concentrations of contaminant which is then rewritten to yield a recursive estimator. Finally this estimator is modified in order to meet the requirements as set forth above.

Let x_i be the daily composite measurement of the concentration of contaminant at day i, $i=1, \ldots, n$. It will be assumed that these are uncorrelated random variables with mean μ_i, the true concentration on day i, and standard deviation σ_i, the known composite measurement accuracy on day i.

One approach to this problem of developing a recursive method of contaminant estimation is to let the estimate of the contaminant be the average of the past values. This estimator has some of the properties listed above. However, it does not have the property of being adaptive to the new data (property a), because after a certain period of time it will settle down and become <u>insensitive</u> to new data. Therefore, a modification to the average is required. Let

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad (4.5.1)$$

denote the arithmetic average of the measured concentrations. Rewrite the above as follows

$$\bar{x}_n = \frac{1}{n}\left[\sum_{i=1}^{n-1} x_i + x_n\right] \tag{4.5.2}$$

Equation (4.5.2) can be written as

$$\bar{x}_n = \left(\frac{n-1}{n}\right) \bar{x}_{n-1} + \left(\frac{1}{n}\right) x_n \tag{4.5.3}$$

Note that:

1. $\frac{n-1}{n} + \frac{1}{n} = 1$; i.e., the sum of the coefficients in (4.5.3) is unity.
2. For large n, $\bar{x}_n \approx \bar{x}_{n-1}$, i.e., the estimate $\bar{x}_n$ becomes insensitive to new data when n is large.

The obvious modification to $\bar{x}_n$ is made using observation (1) above as follows. Denote by $\hat{x}_n$ the modified estimator given by

$$\hat{x}_n = (1-\gamma)\, \hat{x}_{n-1} + \gamma x_n \tag{4.5.4}$$

with

$$\hat{x}_1 = x_1,\ 0 < \gamma < 1. \tag{4.5.5}$$

Observe the following results:

- If γ is near zero the estimator $\hat{x}_n$ is unresponsive to new data.
- If γ is near unity the estimator puts most weight on the last observation.

Now $\hat{x}_n$ can be interpreted as the estimate of a long-term weighted average (over several days) with higher weighting on the most recent data. This interpretation will be made clear later. Thus, $\hat{x}_n$ indicates whether the standard is consistently exceeded.

The estimator $\hat{x}_n$ has all the desired properties outlined at the beginning of this subsection. In implementing it a value of γ must be chosen a priori. It is recommended that a value of $\gamma=.2$ be used.

In order to make confidence statements about the multiday average the distribution of $\hat{x}_n$ will be required. If

$$x_i \sim \mathcal{N}(\mu_i, \sigma_i) \tag{4.5.6}$$

then this implies that the estimate $\hat{x}_n$ is also normal. Namely,

$$\hat{x}_n \sim \mathcal{N}(\hat{\mu}_n, \hat{\sigma}_n^{\,2}) \tag{4.5.7}$$

where

$$\hat{\mu}_n = E\,\hat{x}_n = \sum_{i=2}^{n} \gamma\,(1-\gamma)^{n-i}\,\mu_i + (1-\gamma)^{n-1}\,\mu_1 \tag{4.5.8}$$

and

$$\hat{\sigma}_n^{\,2} = \mathrm{VAR}\,\hat{x}_n \tag{4.5.9}$$

As shown in Appendix A, $\hat{\sigma}_n^{\,2}$ is given recursively by the equation

$$\hat{\sigma}_n^{\,2} = (1-\gamma)^2\,\hat{\sigma}_{n-1}^{\,2} + \gamma^2\sigma_n^{\,2} \tag{4.5.10}$$

Connection with the Discounted Least Squares Approach

The method of discounted least squares involves minimizing a discounted sum of squares as follows:

Let x_i, i=1, ... n, be the set of samples, β be given such that $0 < \beta < 1$ and λ be an unknown constant. The discounted sum of squares with discount rate β is

$$\phi(\lambda) = \sum_{i=1}^{n} (x_i - \lambda)^2\,\beta^{n-1} \tag{4.5.11}$$

Then the method of discounted least squares involves minimizing $\phi(\lambda)$ given above with respect to λ. Let $\hat{\lambda}_n$ be the value that minimizes $\phi(\lambda)$. Then it is well known [K1] that

$$\hat{\lambda}_n = \frac{\sum_{i=1}^{n} \beta^{n-i} x_i}{\sum_{i=1}^{n} \beta^{n-i}} \tag{4.5.12}$$

Appendix B shows that the estimates $\hat{\lambda}_n$ and $\hat{x}_n$ are equivalent for large n if

$$\gamma = 1 - \beta \tag{4.5.13}$$

Equation (4.5.11) can be interpreted as follows:

- Let $x_1, x_2, \ldots, x_n$ be n observed concentrations of contaminants.
- Then $\hat{\lambda}_n$ is an estimate of the average of the $x(i)$'s with weights

  1 on x_n -- the last observation

  β on x_{n-1} -- next to last observation

  $\vdots$

  β^{n-1} on x_1 -- the first observation.

In other words, "older" data points are weighted less and less, i.e., discounted exponentially. Furthermore, since $0 < \beta < 1$, after some point β^k becomes so small that x_{n-k} has negligible weight in the estimates $\hat{\lambda}_n$.

5. THE DECISION ON INSTITUTING CONTROLS

In this Section a procedure is developed for the decision on whether engineering controls should be instituted in an industrial environment. The criterion for the decision is the following: Controls are to be instituted if there is high confidence that the standard is exceeded more than a given percentage of days.

The mathematical formulation of the above criterion is given in Subsection 5.1 in terms of the "probability of violation". The method to estimate this probability of violation and carry out the decision procedure is presented in Subsection 5.2. The estimation of the probability of violation is done using a discounted average, similar to the one presented in the previous section. The main reason for the discounting is the desirability of having a procedure that, while accounting for several days' data, has the capability of following trends. Thus, if a slow increase in contaminant level over days takes place, the procedure will react to it, but if only one "exception" occurs it will not indicate need to institute controls. This is the basic philosophy for instituting controls: If the likelihood of repeatedly having an unsafe environment exceeds a threshold then controls are to be instituted.

5.1 MATHEMATICAL FORMULATION

The best indication of the percentage of days the exposure exceeds the standard in the long run is given by the probability that such an excess can occur on a given day. The contaminant levels on different days are assumed to be independent random variables. Based upon available data one can estimate the probability that the standard will be exceeded in one particular day. Since such an event is equivalent to a violation of the standard, it will be called "probability of violation" and denoted as p_v. It will be assumed that p_v is an unknown constant and a test is to be performed between the following hypotheses

$$H_o: \quad p_v \leq \tilde{p} \qquad (5.1.1)$$

$$H_1: \quad p_v > \tilde{p}$$

where $\tilde{p}$ is the probability threshold for instituting controls. The probability of type I error has to be less than α. Thus, if one rejects H_o at level α then controls should be instituted.

The procedure to carry out the test is by obtaining the LCL (lower confidence limit) with confidence $1-\alpha$ and comparing it to the threshold $\tilde{p}$. If it exceeds $\tilde{p}$ then H_1 is accepted, i.e., controls are to be instituted; otherwise H_o is accepted and no action is to be taken.

5.2 THE CONFIDENCE REGION FOR THE PROBABILITY OF VIOLATION (LONG-TERM SAMPLES)

The procedure will be described next for the case where the daily data on the 8-hr. average have been obtained from long-term samples.

The following notations are used:

x_i - daily composite measurement on day i.

σ_i - standard deviation of this measurement

μ_i - true value of the 8-hr. average exposure on day i.

The relationship between the above variables is

$$x_i = \mu_i + w_i \tag{5.2.1}$$

where the measurement error w_i is a normal, zero-mean random variable with known variance σ^2_i.

If the upper confidence limit, with probability of error α_i, on μ_i is below the standard then a worker is unexposed on the given day (cf. Section 4.1). The upper confidence limit is defined by the equation

$$\int_{-\infty}^{UCL_i} \mathcal{N}(x_i;\ \mu_i,\ \sigma_i^2)\ dx_i = 1 - \alpha_i \tag{5.2.2}$$

where

$$\mathcal{N}(x;\ \mu,\ \sigma^2) = (2\pi)^{-1/2}\ \sigma^{-1}\ \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \tag{5.2.3}$$

and the integration variable has been written out explicitly. It is assumed that the tail of the normal density in the negative region is negligible.

Based upon (5.2.2) we say that there is a confidence $1-\alpha_i$ that $\mu_i < UCL_i$ and a confidence (probability) α_i that $\mu_i > UCL_i$.

Similarly, the confidence that μ_i < STD=1 is

$$\int_{-\infty}^{1} \mathcal{N}(x_i;\ \mu_i,\ \sigma_i^2)\ dx_i \triangleq 1 - p_i \tag{5.2.4}$$

Now p_i is the confidence (probability) that μ_i > STD=1, i.e.

$$\int_{1}^{\infty} \mathcal{N}(x_i;\ \mu_i,\ \sigma_i^2)\ dx = p_i \tag{5.2.5}$$

This will be considered as a "measurement" of the probability of violation based on data from day i. As discussed earlier, the true probability of violation, p_v, is an indication, in the long run, of the percentage of days the exposure exceeds the standard.

Denote the logarithm of the true probability of violation as

$$\pi_v = \log p_v \tag{5.2.6}$$

The "measurement" of π_v on day is

$$\pi_i = \log p_i \tag{5.2.7}$$

and define $\hat{\pi}_n$ as the estimate of π_v. This estimate is taken as the discounted average of π_i, i=1, ..., n, given by the following recursion (see Appendix C)

$$\hat{\pi}_n = (1 - \gamma)\ \hat{\pi}_{n-1} + \gamma\pi_n \tag{5.2.8}$$

The "gain" of the above recursive estimator of π_v is

$$\gamma = 1 - \beta \tag{5.2.9}$$

where β is the discounting factor. As in Subsection 4.5, the choice of $\gamma=.2$ is recommended. This has an effective memory length of approximately 5 days. For $n \leq 5$ a uniform weighting is to be used, namely

$$\hat{\pi}_n = \frac{1}{n} \sum_{i=1}^{n} \pi_i , \qquad n \leq 5 \tag{5.2.10}$$

while for $n>5$, (5.2.8) is to be used.

The reasons for estimating the probability of violation via a discounted average are the same as those mentioned in Subsection 4.5. The most important of them is the adaptivity feature.

The sample standard deviation associated with the estimate (5.2.8) can be obtained using the recursion (see Appendix C)

$$\hat{\tau}_n^2 = (1 - \gamma)\, \hat{\tau}_{n-1}^2 + (\pi_n - \hat{\pi}_n)^2 , \; n > 5 \tag{5.2.11}$$

and

$$\hat{\tau}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\pi_i - \hat{\pi}_n)^2 , \; n \leq 5 \tag{5.2.12}$$

is the start-up procedure.

For the purpose of the test of the hypotheses (5.1.1), the distribution of $\hat{\pi}_n$ is needed. The use of the beta densities, which are defined on the interval [0,1], is standard for computing confidence intervals for probabilities [K1]. However, in the present case, the probabilities p_i will be mostly concentrated around zero (small probability of violation) and a good approximation of the beta density around zero is a lognormal density. This will also yield a simple procedure that meets the requirements of minimum computational burden on the employer. Therefore, π_i, defined in (5.2.7), will be normally distributed and their random unknown mean is assumed to be π_v, the logarithm of true probability of violation. Then the hypothesis testing can be performed as follows.

The LCL on π_v is

$$LCL(\pi_v) = \hat{\pi}_n - t\,\hat{\tau}_n \qquad (5.2.13)$$

where t is obtained from a t-distribution table for the desired confidence.

Thus, the test whether p_v exceeds threshold $\tilde{p}$ can be carried out by testing whether $\pi_v = \log p_v$ exceeds $\tilde{\pi} = \log \tilde{p}$ using the expression of (5.2.13).

5.3 THE CONFIDENCE REGION FOR THE PROBABILITY OF VIOLATION (GRAB SAMPLES)

Following [B1] the estimate of the 8-hr. exposure will be taken as the composite measurement, i.e.,

$$x_i = 10^{\bar{y} + \frac{1}{2}a\, s^2} \qquad (5.3.1)*$$

where, as in [B1], $\bar{y}$ and s are the sample mean and standard deviation of the common logs of the measurements on day i normalized w.r.t. the 8-hr. standard (no subscripts are attached here to $\bar{y}$ and s for simplicity) and $a = \ln 10$. The true exposure for that day will be assumed approximately normally distributed about x_i with variance to be derived next.

Let $\bar{y}_T$ and σ_T^2 be the true values estimated by $\bar{y}$ and s^2, respectively. Then the true 8-hr. exposure is

$$m_i = 10^{\bar{y}_T + \frac{1}{2}a\, \sigma_T^2} \qquad (5.3.2)$$

---

*The correction factors used in [B1] are close to unity and thus are neglected.

Thus,

$$\sigma_i^2 = E(x_i - m_i)^2$$

$$= E\left[10^{\bar{y} + \frac{1}{2}a\, s^2} - 10^{\bar{y}_T + \frac{1}{2}a\, \sigma_T^2}\right]^2$$

$$= \left[10^{\bar{y}_T + \frac{1}{2}a\, \sigma_T^2}\right]^2 E\left[1 - 10^{\bar{y} - \bar{y}_T + \frac{1}{2}a(s^2 - \sigma_T^2)}\right]^2$$

$$= m_i^2\, E\left[1 - e^{a(\bar{y} - \bar{y}_T) + \frac{1}{2}a^2(s^2 - \sigma_T^2)}\right]^2 \qquad (5.3.3)$$

Using a second order expansion of the term in the brackets above yields

$$\sigma_i^2 = m_i^2\left[\frac{1}{2}a^2\, E(\bar{y} - \bar{y}_T)^2 + \frac{1}{8}a^4\, E(s^2 - \sigma_T^2)^2\right] \qquad (5.3.4)$$

where the first order terms are zero because

$$E\bar{y} = \bar{y}_T \quad , \qquad Es^2 = \sigma_T^2 \qquad (5.3.5)$$

If the number of samples on day i was n_i then (5.3.4) can be approximately written as

$$\sigma_i^2 = x_i^2\left[\frac{a^2 s^2}{2\, n_i} + \frac{a^4 s^4}{4n_i}\right]$$

$$= x_i^2\, \frac{2a^2\, s^2 + a^4\, s^4}{4n_i} \qquad (5.3.6)$$

where n_i is the number of short-term samples on day i.

While the normality assumption is only an approximation, there are several justifications for it.

a. The exact analysis would not be implementable without the use of a scientific computer.

b. Most of the contaminants fall in the category treated in [L1] and it is desirable to have a unified procedure.

c. Asymptotically the estimate (5.3.1) is normally distributed.

6. RECOMMENDED INTERVALS FOR SAMPLING

In this section the guidelines for the choice of sampling intervals (number of days after which new measurements are to be made) are presented.

The sampling interval is a function of the observed levels of exposure Table 6.1 presents the legal requirements on sampling [B2] as well as the recommended ones. The recommended sampling interval for unexposed workers is six months (120 working days) while for overexposed workers is one week (5 days). For the intermediate class, i.e., exposed, a systematic procedure to obtain it is presented next.

Classification Of Employee	Minimum Sampling Frequency	Recommended Sampling Frequency
A. Unexposed	None	Once every six months
B. Exposed	Once every 2 months	According to Fig. 6.1
C. Overexposed	Once every 1 month	Once every week

Table 6.1. Sampling Frequencies

The decision on the length of the sampling intervals is to be based on the (estimated) percentage of days with exposures exceeding the standard. As discussed in Section 5, the best estimate for this is the "probability of violation". This probability is obtained using the sequential procedure described in Section 5.

The rule for obtaining the sampling interval has to satisfy the following predetermined constraints:

a. For a daily composite mean x_i=.5 (i.e. 1/2 standard) and a geometric standard deviation GSD=1.3 the sampling interval should be 90 days (3 months).

b. For $x_i = .8$ with the same GSD the interval should be 7 days (1 week).

The probability of violation can be calculated, using the fiducial approach (which is entirely equivalent in this case to the confidence region method), from the following expression

$$p_v = \int_1^{\infty} \mathcal{N}(x \; ; \; x_i, \sigma_i) \, dx \qquad (6.1)$$

where σ_i is the standard deviation obtained as $(GSD-1)x_i$. For the above cases the values of σ result as .15 and .24, respectively. The corresponding values of p_v are then 4×10^{-4} and .2, respectively.

Figure 6.1 presents, on a semilog scale, a straight line interpolation between the above two points. For $p_v > .2$ the sampling interval is 5 days while for $p_v < 4 \times 10^{-4}$ it increases to the maximum of 120 days.

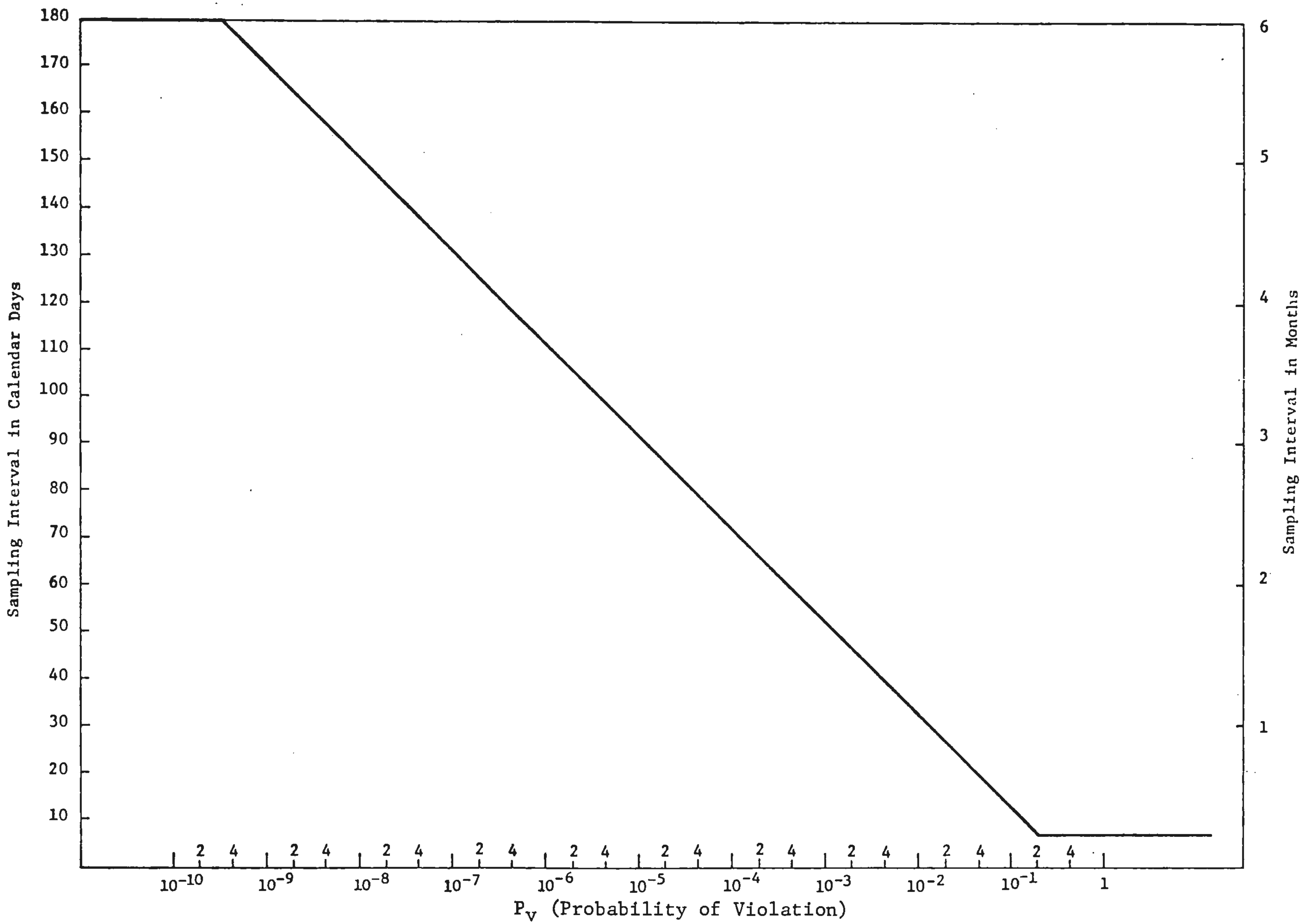

Figure 6.1 Calculation of Sampling Interval

7. SUMMARY AND CONCLUSION

A sequential procedure to be used by employers to ensure the safety of their employees in an industrial environment with atmospheric contaminants has been developed. This procedure, which is documented in a user-oriented fashion in a companion Handbook [B2], has been developed with the intent of minimizing the sampling as well as computational burden on the employer. The various steps of which it consists have been identified and the corresponding subprocedures have been developed in such a way that they are field-implementable, i.e., with modest computational requirements.

Further investigations on the statistical properties of the atmospheric contaminants should be made in the future in order to increase the confidence that correct statistical models are being used for the various decision procedures.

APPENDIX A

THE VARIANCE OF THE RECURSIVE ESTIMATE OF THE MULTIDAY EXPOSURE

Let x_i, $i=1, \ldots, n$ be as in (4.5.6) and $\hat{x}_n$ given by (4.5.4). Then it is required to show that

$$\hat{x}_n \sim \mathcal{N}(\hat{\mu}_n, \hat{\sigma}_n^2) \tag{A.1}$$

where

$$\hat{\mu}_n = E\,\hat{x}_n = \sum_{i=2}^{n} \gamma(1-\gamma)^{n-i} \mu_i + (1-\gamma)^{n-1} \mu_1 \tag{A.2}$$

$$\hat{\sigma}_n^2 = (1-\gamma)^2 \hat{\sigma}_{n-1}^2 + \gamma^2 \sigma_n^2 \;. \tag{A.3}$$

In order to demonstrate (A.1) it is sufficient to write $\hat{x}_n$ explicitly in terms of $x(t_1), \ldots, x(t_n)$. Solving (4.5.4) by backward substitution, the following result is obtained:

$$\hat{x}_n = \sum_{i=2}^{n} \gamma(1-\gamma)^{n-i} x_i + (1-\gamma)^{n-1} x_1 \tag{A.4}$$

The fact that $\hat{x}_n$ is normally distributed now follows because it is a linear combination of normally distributed random variables. Furthermore, it immediately follows that $E\hat{x}_n$ is as given by equation (A.2). To obtain an iterative version of the variance of $\hat{x}_n$, note that

$$\hat{x}_n - \hat{\mu}_n = (1-\gamma)\,(\hat{x}_{n-1} - \hat{\mu}_{n-1}) + \gamma(x_n - \mu_n)$$

and thus

$$\text{Var}\,\hat{x}_n = (1-\gamma)^2\,\text{Var}\,\hat{x}_{n-1} + \gamma^2\,\text{Var}\,x_n \tag{A.5}$$

follows because x_n is independent of $\hat{x}_{n-1}$ by assumption. This completes the proof of (A.3).

APPENDIX B

EQUIVALENCE OF THE RECURSIVE ESTIMATE WITH THE DISCOUNTED LEAST SQUARES

Let $\hat{x}_n$ be as in (4.5.4) and let $\hat{\lambda}_n$ be as in (4.5.12). First note that the sum of the coefficients multiplying the x_i's in both $\hat{x}_n$ and in $\hat{\lambda}_n$ are equal to one.

The coefficients multiplying the x_i's in $\hat{x}_n$ are*

$$\delta_i' = \gamma(1-\gamma)^{n-i} \qquad i=2,\ldots, n \tag{B.1}$$

The coefficients multiplying the x_i's in $\hat{\lambda}_n$ are:

$$\delta_i'' = \frac{\beta^{n-i}}{\sum_{i=1}^{n} \beta^{n-i}} = \frac{\beta^{n-i}(1-\beta)}{1-\beta^n} \tag{B.2}$$

where the summation formula for a geometric series has been used. For $n\to\infty$, but $n-i$ finite, one has

$$\delta_i'' = \frac{\beta^{n-i}(1-\beta)}{1-\beta^n} = \beta^{n-i}(1-\beta) \tag{B.3}$$

Therefore, for large n if

$$\gamma = 1-\beta \tag{B.4}$$

it follows from (B.1) and (B.3) that

$$\delta_i' = \delta_i'' \tag{B.5}$$

for i "close" to n. This completes the proof of equivalence of the two estimates $\hat{x}_n$ and $\hat{\lambda}_n$.

---

*Asymptotic properties are being considered and thus $\delta_1' = (1-\gamma)^n$ can be neglected since it becomes zero.

APPENDIX C

THE ESTIMATE OF THE PROBABILITY OF VIOLATION AND ITS VARIANCE

The discounted estimate of the logrithm of the probability of violation is given by the following equation

$$\hat{\pi}_n = c_n^{-1} \sum_{i=1}^{n} \beta^{n-i} \pi_i \tag{C.1}$$

where β is the discount rate and

$$c_n = \sum_{i=1}^{n} \beta^{n-i} \tag{C.2}$$

As discussed in Appendix A, this is equivalent to the recursive form (5.2.8) i.e.,

$$\hat{\pi}_n = (1 - \gamma) \hat{\pi}_{n-1} + \gamma\pi_n \tag{C.3}$$

The sample variance of $\hat{\pi}_n$ is, for the case of uniform weighting (no discounting)

$$\hat{\tau}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\pi_i - \hat{\pi}_n)^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} (\pi_i - \hat{\pi}_n)^2 + \frac{1}{n-1} (\pi_n - \hat{\pi}_n)^2 \tag{C.4}$$

Assuming that $\hat{\pi}_{n-1} \approx \hat{\pi}_n$ for the purpose of calculating the variance, the above becomes

$$\begin{aligned} \hat{\tau}_n^2 &= \frac{1}{n-1} \sum_{i=1}^{n-1} (\pi_i - \hat{\pi}_{n-1})^2 + \frac{1}{n-1} (\pi_n - \hat{\pi}_n)^2 \\ &= (1 - \frac{1}{n-1}) \hat{\tau}_{n-1}^2 + \frac{1}{n-1} (\pi_n - \hat{\pi}_n)^2 \end{aligned} \tag{C.5}$$

or, the "fixed gain" case, (i.e., discounted estimate) as given by (C.3) with gain γ it becomes

$$\hat{\tau}_n^2 = (1 - \gamma)\ \hat{\tau}_{n-1}^2 + \gamma\ (\pi_n - \hat{\pi}_n)^2 \tag{C.6}$$

The start-up procedure for the variance for $n \leq 5$ is using the following equal-weighting expression

$$\hat{\tau}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\pi_i - \hat{\pi}_n)^2 \qquad n \leq 5 \tag{C.7}$$

and then, for $n > 5$, (C.6) is to be used.

APPENDIX D

THE BETA DENSITY

The normalized beta density function is defined as:

$$f_B\ (z|r,\ n) = \frac{1}{B(r,\ n-r)} \cdot z^{r-1}(1 - z)^{n-r-1} \tag{D.1}$$

on the domain [0, 1], i.e.,

$$0 \leq z \leq 1 \tag{D.2}$$

and with parameters

$$n > r > 0. \tag{D. 3}$$

The term B(r, n-r) is the normalization constant. It can be shown that if x has a beta distribution then [K1]

$$E(x) = \frac{r}{n} \tag{D.4}$$

$$\text{var}(x) = \frac{r(r+1)}{n(n+1)} \tag{D.5}$$

Figure D.1 illustrates two beta densities for E(x) = .05 and var (x) = .0028 and var(x) = .0033 respectively. An examination of this figure illustrates the similarity between these beta densities and a "narrow" lognormal density.

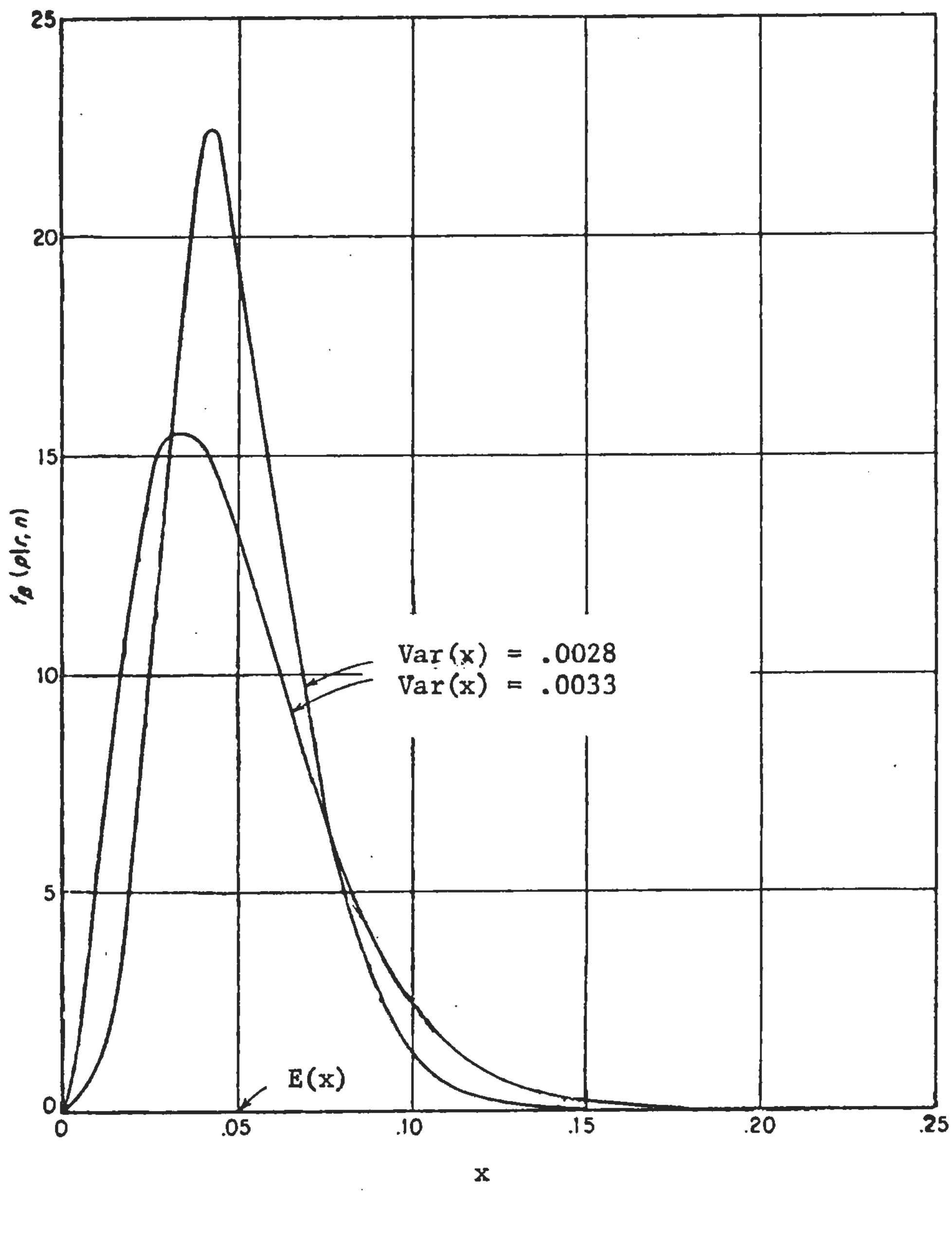

FIGURE D.1 EXAMPLE OF BETA DENSITIES

REFERENCES

[B1] Y. Bar-Shalom, et. al. "Statistical Theory for Sampling of Time-Varying Industrial Atmosphere Contaminant Levels," Final Report, Systems Control, Inc., Palo Alto, California, June 1974.

[B2] D. H. Budenaers and Y. Bar-Shalom, "Handbook for Sequential Sampling Plans and Decision Theory for Employer Monitoring of Employee Exposure to Industrial Atmospheres," Systems Control, Inc., Palo Alto, California, May 1975.

[D1] A. J. Duncan, Quality Control and Industrial Statistics, R. D. Irwin, Inc., 1965.

[F1] The Occupational Safety & Health Act of 1970 (PL91-596); Standards in the 18 Oct. 1972 Federal Register (29 CFR 1910.93).

[K1] Kendall, M. S. and Stuart, A.: The Advanced Theory of Statistics, Hafner Publishing Co., New York, Vol. I, (1969), Vol. II, (1967).

[L1] Leidel, N. A. and Busch, K. A.: "Statistical Methods for the Determination of Noncompliance with Occupational Health Standards," TR-76, U. U. S. Department of Health, Education, and Welfare, Public Health Service, National Institute for Occupational Health, Division of L Laboratories and Criteria Development, Cincinnati, Ohio, (1976).

[P1] E. Parzen, Modern Probability Theory and Its Application, Wiley, 1960.