Benzene exposure in the petroleum refining industry

ROBERT C. SPEAR, STEVE SELVIN, JANE SCHULMAN and MARCIE FRANCIS
Northern California Occupational Health Center, School of Public Health: University of California, Berkeley, California 94720

Benzene exposure data, Submitted by nine petroleum refining companies, were studied with the general objective of characterizing the distribution of exposures within work operations and job categories. The data were collected by company hygienists over the period of years between 1978 and 1984. All measurements were of personal exposures determined using charcoal tubes or organic vapor monitors. Of the 123 location and unit specific job groups studied, most eight-hour time-weighted average (TWA) exposures were below 1.0 ppm. For some groups, however, the variability in exposure was such that ten percent or more of exposures exceeded 1.0 ppm. For nineteen of the most highly exposed groups, sufficient data were available to study the variability in exposure associated with individual workers versus the common work environment. Results indicate that there were some cases in which variability in exposure was mostly associated with the environment and others in which it was mostly associated with differences between workers, but it was most often approximately evenly split between the two sources. This analysis provides guidance in determining whether control strategies should be targeted at the work environment or at work requirements or practices of individual workers within the group. Some short-term exposure data were included for study. It was found that most 15-minute TWA exposures were less than 1.0 ppm, but again, there were groups with highly variable exposure with some measurements in excess of 5 ppm. These were often in jobs involving loading and unloading operations of barges or tanker trucks. **Spear, R.C.; Selvin, S.; Schulman, J.; Francis, M.:** Benzene exposure in the petroleum refining industry. *Appl. Ind. Hyg. 2:*155–163; 1987.

Introduction

In the latter part of 1984, the authors were asked by the American Petroleum Institute to conduct a study of benzene exposure data available from some of their member companies. These data had been collected by company hygienists over the years 1978 to 1984. The goals of the study were to:

1. Characterize the distribution of exposures within work operations and job categories.
2. Estimate the proportions of individuals exposed to benzene at or above levels ranging from 0.5 to 2.0 ppm.
3. Investigate the usefulness and suitability of specific statistical distributions as summary descriptors of exposure, e.g., normal, log-normal, etc.
4. Investigate intrapersonal variation as a proportion of total variation in exposure levels within job categories.
5. Assess the implications of the results as they relate to air sampling strategies and the definition of compliance with OSHA standards.

These goals can be broadly classified into descriptive, analytical, and policy-related aspects of the study. This paper deals first with a description of the benzene exposure levels in various parts of the industry during the time period in question. An analytical framework is presented which was found to be necessary to characterize exposures adequately and to deal with issues of exposure variability within and between workers. The policy implications of these analyses, in the context of item 5 above, will be treated in a subsequent paper.

The data base

The data base of personal exposure measurements which was available for analysis was comprised principally of eight-hour time-weighted average (TWA) values, although some short-term exposure measurements were submitted. An early decision was made to focus on the eight-hour data to avoid the additional variability that would be introduced by different sampling durations.[1] All measurements were collected with either charcoal tubes or passive organic vapor monitors.

Data were submitted by nine petroleum refining companies. The measurements were originally collected for many different reasons by the company hygienists; for the majority of the data, it was impractical to consider separately those samples collected under routine conditions versus those collected to characterize exposures during unusual circumstances. In that regard, respiratory protection was used in some of these exposure situations although data on the frequency of respirator use were not analyzed. In general, each measurement was classified by worker identifier, location, unit, and job. The fact that more than one observation was

0882-8032/87/040155-09$2.50 © ACGIH 1987

collected on most workers presented the possibility of estimating the variability in exposure due to the common work environment versus that associated with a particular individual.

Early in the study, considerable effort was given to working with the company hygienists in an attempt to define job categories that would at least approximate equal exposure groups, i.e., groups of workers whose exposures could be expected to be similar based on the nature and location of their jobs. In some companies, the specificity of job titles made it necessary to pool several job categories into a single exposure group for analysis. In the end, all job groups were specific to a particular refinery, a unit within the refinery, and an operation within the unit. For example, one particular job group was comprised of all benzene production operators (24 persons and 58 measurements) on the benzene and cyclohexane unit at location B.

Descriptive results: eight-hour TWA measurements

Once an understanding had been reached on how to specify job groups, the companies were asked to submit eight-hour exposure data for all such groups for which there were 30 or more measurements collected from ten or more workers. There were 123 groups which met these requirements approximately, with some having as few as eight workers. These data are described by giving the percent of the measurements below the limits of detection and the fraction of measurements observed above 0.5, 1.0, and 2.0 ppm; these are values of potential regulatory significance.

Table I contains the descriptive exposure data for groups where at least ten percent of measurements exceeded 0.5 ppm, i.e., data are reported for the most highly exposed groups. The columns of Table I, labeled $P(X>0.5)$, $P(X>1.0)$, and $P(X>2.0)$, are the fractions of the total number of measurements above 0.5, 1.0, and 2.0 ppm, respectively. These values are referred to below as exceedance fractions. The limits of detection vary by sampling medium and by analytical method. Since these factors differed from company to company and over time within companies, detection limits were reported separately for each data submission.

For 30 of the 123 groups, all of the measurements were below 0.5 ppm. For many of these 123 groups, the fraction below the limit of detection was also quite high. On the other end of the scale, roughly 15 percent of these groups have significant fractions of the measurements in excess of 1.0 ppm. The groups with relatively high exposures and with adequate numbers of measurements will be considered below in some detail. The high exposure groups are comprised of operators of production units within the refineries, bulk transfer or loading operations, and laboratory quality control activities.

There was some interest in the exposure of maintenance workers, and a special effort was made to obtain data on this employee group. The data submitted were eight-hour TWA values and are summarized in Table II. However, the nature of maintenance work is such that these groups are not unit specific. Most, in fact, are only company specific so that there can be no pretense of uniformity of exposure. Nevertheless, it should be noted that only about five percent of the measurements exceed 1.0 pprn, and a large percentage of the measurements are below detection limits.

Descriptive results: short-term exposures

An effort was made, late in the study, to obtain data on short-term benzene exposures. In general, short-term refers to exposures measured over intervals of 5 to 30 minutes when a particular task or work condition leads the hygienist to expect high transient exposures. The format for data collection and analysis was, as with the eight-hour data, focused on unit and site specific job groups. In this case, we sought at least ten measurements per group, but no further constraints were imposed. The majority of data was submitted by one company, but there are some data from a total of four companies.

Table III contains a summary of the data in the same format as was used for the eight-hour TWA data. The exceedance fractions are based only on a count of the number of measurements above each of the limits where each measurement is treated as an independent value. In all cases, the sampling medium was charcoal tubes. Because the sampling time was not uniform, the values were normalized to a 15-minute standard. If the measurement duration was less than 15 minutes, it was assumed, following OSHA compliance practice, that the hygienist making the measurement had concluded that the task was complete and no further exposure would have occurred. Hence, the total mass collected would not have increased had sampling continued for a longer period, and the TWA value was thereby adjusted downward by the ratio of the sampling duration to the 15-minute standard. If the duration exceeded 15 minutes, the TWA value was left unaltered under the assumption that exposure was uniform over time, the only possible choice in the absence of other information.

As can be seen from Table 111, most of the short-term exposures are very low. However, as with the eight-hour data, there are some groups that show relatively high exceedance fractions even at 5 ppm. Those groups showing exceedance fractions of 0.1 or above at 5 pprn are involved in tank gauging, loading and unloading operations, barge transfer operations, and tanker truck loading.

For some groups in Table III, groups 29 to 31 for example, the limits of detection were relatively high. This leads to high proportions below the detection limit, but also high proportions of the measurements in excess of 0.5 ppm. These entries are not in error and simply indicate short sample duration and correspondingly high detection limit situations.

Analytical framework

The descriptive data leaves one with the impression that, to the extent that these data are representative of the industry, most operations are associated with exposures below 1.0 pprn most of the time. Some operations, however, have significant fractions of their exposures above 1.0 ppm. To reach any more specific conclusion on the basis of these descriptive data alone is difficult. Indeed, the second goal of the study, which was to estimate the proportions of ***individuals*** who were exposed to various benzene levels, cannot be addressed since these data relate to the proportion of measurements rather than of individuals. What is needed is a framework to disentangle the individual worker's contribution to overall variability from those environmental factors that contribute to the variability in exposure common to the group.

Oldham and Roach were apparently the first to apply analysis of variance

TABLE I

Empiric estimates of exceeding certain limits (ppm) for groups of refinery workers exposed to benzene

Group	Description	Medium[A]	Samples	Persons	P(X>0.5)	P(X>1.0)	P(X>2.0)	% BLD[B]
1	Location A, Catalytic Cracker Refining Operators	OV	202	39	0.01	0.00	0.00	27.7
2	Location B, Ethylene Unit Benzene Production Operators	OV	170	44	0.09	0.01	0.01	8.2
3	Location B, Waste Water Treatment Refining Operators	OV	83	22	0.37	0.13	0.06	15.7
4	Location B, Benz. & Cyclohex. Unit Benzene Production Operators	OV	58	24	0.41	0.14	0.03	0.0
5	Location E, CRU Platformer Head Operator, Controlman	CT	55	16	0.13	0.09	0.05	3.6
6	Location E, CRU Platformer Misc. Operators	CT	66	22	0.20	0.17	0.14	7.6
7	Location E, Petrochemical Plant Head Operator, Controlman	CT	80	20	0.28	0.20	0.10	2.5
8	Location E, petrochemical Plant Misc. Operators	CT	161	38	0.35	0.25	0.17	3.1
9	Location F, Benzene Extraction Unit Operator	OV	124	23	0.26	0.12	0.03	10.5
10	Location F, Aromatics-East ACU Operator	OV	97	25	0.20	0.11	0.07	18.6
11	Location F, Aromatics Gauger/Operator	OV	113	33	0.16	0.04	0.00	15.9
12	Location F, Aromatics Column Operator	OV	98	25	0.24	0.07	0.01	14.3
13	Location F, Aromatics-West ACU Operator	OV	79	19	0.11	0.06	0.01	25.3
14	Location F, Aromatics Foreman	OV	89	17	0.24	0.11	0.03	14.6
15	Location F, Cumene-Phenol/Acetone Operator	OV	177	25	0.24	0.12	0.08	24.3
16	Location A, Cracking Department Still-Platformer, Refining Oper.	OV	90	18	0.10	0.02	0.00	2.2
17	Location A, Petrochemical Oept., Cumene/Benzene, Benz. Prod. Oper.	OV	59	20	0.49	0.19	0.05	0.0
18	Location S, Benzene Transfer/Movement Operators	CT	48	24	0.33	0.25	0.19	6.2
19	Location T, Lab Unit Technical	CT	55	18	0.23	0.14	0.02	9.1
20	Location G, Main Deck, Transportation-Marine	CT	49	n/a[c]	0.61	0.55	0.41	0.0
21	Location B, Bulk Oil Pumphouse Refining Operators	OV	33	14	0.36	0.12	0.00	15.2
22	Location B, Bulk Oil Pump Station Refining Operators	OV	45	16	0.18	0.13	0.07	26.1
23	Location B, Laboratory Staff	OV	37	34	0.35	0.08	0.03	0.0
24	Location E, Petrochemical Unit	CT	96	56	0.11	0.08	0.05	14.6
25	Location M, Lab QC Unit All Staff	CT	38	32	0.37	0.21	0.13	2.6
26	Location F, Dispatching Dock Dockman	OV	45	23	0.36	0.22	0.13	26.7
27	Location F, Phenol/Acetone Foreman	OV	116	19	0.14	0.08	0.04	49.1
28	Location F, Oil Recovery-Environmental, Operator	OV	35	22	0.11	0.06	0.03	42.9
29	Location Q, Platformer, Hydro-desulfurizer, Stillman, Refining	CT	56	10	0.14	0.00	0.00	73.2
30	Location Q, Unit D Stillman, Refining	CT	42	17	0.14	0.05	0.00	64.3
31	Location Q, Control Room Laborer, Yardman, Roustabout	CT	36	a	0.11	0.00	0.00	0.0
32	Location R, Compound Plant Yardman	CT	36	8	0.11	0.00	0.00	0.0
33	Location U, Maintenance	CT	37	34	0.14	0.03	0.03	16.2
34	Location U, Lab, Quality Control	CT	52	47	0.19	0.10	0.08	19.2
35	Location W, Receipt. Storage, and Movement	CT	55	36	0.49	0.25	0.20	10.9
36	Location X, Lab Unit, Technical	CT	35	17	0.11	0.03	0.00	11.4
37	Location Y, Catalytic Cracking Unit Outside Journeyman	CT	34	22	0.12	0.09	0.09	11.8
38	Location Y, Aromatics Recovery Unit, Outside Journeyman	CT	34	18	0.50	0.26	0.09	0.0
39	Location AA, Benz. Unit Operators	CT	58	39	0.31	0.16	0.03	6.9
40	Location AA, Ethylene Unit Feedstock Operators	CT	30	24	0.10	0.10	0.00	133
41	Location BB, Lab Tech	CT	46	32	0.11	0.00	0.00	41.3
42	Location CC, Lab Tech	CT	48	12	0.60	0.50	0.27	20.8
43	Location DD, Lab Tech	CT	106	63	0.21	0.07	0.03	8.5
44	Location II, Lab Tech	CT	44	20	0.11	0.02	0.00	52.3

[A]Medium, OV = Passive Organic Vapor Monitor. CT = Charcoal Tube
[B]BLD = Below the limit of detection
[C]n/a = not applicable since there are no ID numbers

TABLE II

Empiric estimates of exceeding certain limits (ppm) for groups of maintenance workers

Group	Description	Medium[A]	Samples	Persons	P(X>1.0)	P(X>2.0)	% BLD[B]
M1	All Locations Pipefitter/Welder, Pipefitter/Plumber, Welder	CT	133	34	0.00	0.00	11.3
M2	Location O Contract Workers, Mostly Maintenance	OV	138	n/a[C]	0.16	0.06	22.5
[illegible]	Location O Maintenance, Misc.	OV	56	34	0.05	0.00	58.9
M4	Location P Contract Workers, Misc. Maintenance	OV	70	n/a[C]	0.04	0.01	78.6
M5	Location F Pipefitter-Phenol/Acetone-Maintenance	OV	104	47	0.03	0.02	56.7
M6	All Locations Maintenance Unit	CT	58	49	0.09	0.07	51.7
M7	Location B Instrument Shop, Machine Shop, Paint Shop	OV	53	16	0.00	0.00	45.3
M8	All Locations Maintenance Operations	OV	56	40	0.09	0.09	41.1
M9	All Locations Mechanic-General Plant	CT	118	36	0.02	0.00	29.7
M10	All Locations Maintenance-Leadman	CT	52	20	0.04	0.04	59.6

[A]Medium, OV = Passive Organic Vapor Monitor, CT = Charcoal Tube.
[B]BLD = Below the Limit of Detection.
[C]n/a = not applicable since there are no unique ID #'s in these groups.

TABLE III

Probability of exceeding selected limits (ppm) for short-term samples

Group	Description	Medium[A]	Samples	Persons	0.5	1.0	[illegible]	5.0	10.0	%BLD[B]
1	Company 1, Location 2, RS&M[C] Loading-Routine	CT	12	2	0.17	0.08	0.00	0.00	0.00	17
2	Company 1, Location 11, RS&M Unloading-Routine	CT	19	3	0.11	0.11	0.05	0.05	0.00	32
3	Company 1, Location 11, RS&M Manually Gauging Tanks-Routine	CT	11	3	0.27	0.18	0.00	0.00	0.00	9
4	Company 1, Location 12, RS&M Loading-Routine	CT	16	4	0.38	0.13	0.00	0.00	0.00	13
5	Company 1, Location 12, RS&M Unloading-Routine	CT	20	3	0.15	0.00	0.00	0.00	0.00	10
6	Company 1, Location 13, RS&M Loading-Routine	CT	15	3	0.87	0.47	0.27	0.00	0.00	7
7	Company 1, Location 13, RS&M Unloading-Routine	CT	10	3	0.20	0.10	0.10	0.10	0.00	10
8	Company 1, Location 14, RS&M Loading-Routine	CT	12	3	0.50	0.17	0.00	0.00	0.00	0
9	Company 1, Location 14, RS&M Unloading-Routine	CT	14	3	0.29	0.14	0.07	0.00	0.00	14
10	Company 1, Location 15, RS&M Unloading-Routine	CT	24	3	0.33	0.17	0.00	0.00	0.00	13
11	Company 1, Location 16, RS&M Loading-Routine	CT	13	2	0.15	0.00	0.00	0.00	0.00	0
12	Company 1, Location 17, RS&M Loading-Routine	CT	10	3	0.20	0.10	0.00	0.00	0.00	10

TABLE III (cont.)
Probability of exceeding selected limits (ppm) for short-term samples

Group	Description	Medium[A]	Samples	Persons	0.5	1.0	2.0	5.0	10.0	% BLD[B]
13	Company 1, Location 17, RS&M Unloading-Routine	CT	13	2	0.15	0.00	0.00	0.00	0.00	23
14	Company 1, Location 18, RS&M Unloading-Routine	CT	19	3	0.21	0.11	0.00	0.00	0.00	0
15	Company 1, Location 18, RS&M Manually Gauging Tanks-Routine	CT	14	4	0.64	0.64	0.50	0.21	0.00	7
16	Company 1, Location 22, RS&M Unloading-Routine	CT	11	6	0.18	0.00	0.00	0.00	0.00	36
17	Company 1, Location 24, RS&M Loading-Routine	CT	10	3	0.90	0.90	0.70	0.20	0.10	10
18	Company 1, Location 25, RS&M Loading-Routine	CT	44	11	0.30	0.16	0.05	0.00	0.00	25
19	Company 1, Location 26, Petrochemical Plant Area Visual Inspection of Automatic Gauges-Routine	CT	27	4	0.26	0.07	0.00	0.00	0.00	11
20	Company 1, Location 27 Laboratory, Quality Control Testing-Routine	CT	15	9	0.33	0.00	0.00	0.00	0.00	33
21	Company 1, Location 28, RS&M Manually Gauging Tanks-Routine	CT	10	5	0.30	0.10	0.00	0.00	0.00	0
22	Company 3, Location 1 Barge Transfer Operations	CT	18	9	0.94	0.94	0.89	0.72	0.56	0
23	Company 6, Location 1 Tanker Truck Loading-Bottom Without Recovery	CT	18	10	0.22	0.17	0.17	0.11	0.00	67
24	Company 6, Location 2 Tanker Truck Loading-Bottom With Recovery	CT	12	9	0.25	0.08	0.00	0.00	0.00	67
25	Company 6, Location 4 Tanker Truck Loading-Top Without Recovery	CT	10	10	0.40	0.20	0.20	0.10	0.00	40
26	Company 2, Location 6 Crude Processing Unit Outside Journeyman-Routine	CT	14	7	0.14	0.07	0.07	0.07	0.07	0
27	Company 2, Location 7 ARO Recovery Unit Outside Journeyman-Routine	CT	26	9	0.65	0.54	0.35	0.04	0.04	0
28	Company 2, Location 7 Lab Unit Technician-Routine	CT	16	7	0.19	0.06	0.06	0.00	0.00	37
29	Company 2 Marketing Ops-RS&M Admin/Sup-Routine	CT	18	6	0.44	0.33	0.17	0.00	0.00	61
30	Company 2 Marketing Ops-RS&M Terminal Operators-Routine	CT	68	22	0.62	0.26	0.04	0.01	0.01	79
31	Company 2 Marketing Ops-RS&M Transport Driver-Routine	CT	98	46	0.78	0.42	0.11	0.02	0.00	58

[A]MEDIUM, CT = Charcoal Tubes, OV = Passive Organic Vapor Monitors.
[B]BLD = Below the Limit of Detection.
[C]RS&M = Receipt, Storage and Movement.

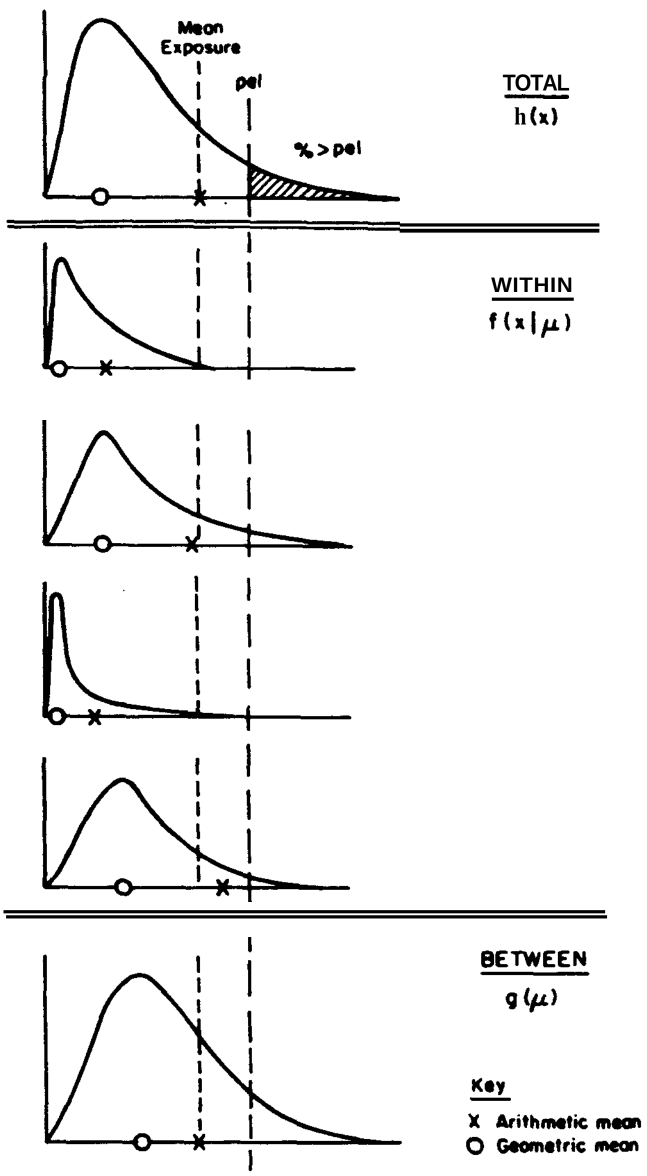

Figure 1—**Lognormal** exposure model.

techniques to exposure data to investigate worker-to-worker versus day-to-day variability in exposure.[2] The concept is shown schematically in Figure 1. The idea is that each worker's exposure differs from day to day and that, over time, the exposure is best described by a distribution of eight-hour TWA measurements which will generally be different for each worker. These are the distributions shown in the middle of the figure. If these distributions are not the same for each worker, then one's view of the exposure of the group, based on a sample of exposure data of the sort presented above, obviously depends upon which workers were selected for measurement and how many times each was measured, hence the difficulty in interpreting the benzene data without directly addressing this issue.

An analysis of variance model was adopted based on the proposition that the distribution of individual exposures was lognormal and that the variance of each of these individual distributions is the same for each worker, but with differing mean levels of exposure. Because the individual variances are assumed identical, the only remaining element of the model to be specified concerns the distribution of mean values among workers. It is not unreasonable to postulate that this distribution is also lognormal. This is the distribution shown at the bottom of the figure. That is, if one were to determine the mean value of each worker's exposure distribution and if there were a large number of workers, the distribution of means would also be lognormal. (The means referred to here are the statistical expectations, estimated by the arithmetic means, of the lognormal distributions; they are not the geometric means or the means of the logarithms. This is a subtle, but important, point which can easily lead to confusion.) Consequently, the exposure distribution obtained by randomly selecting workers and randomly selecting days on which measurements were to be taken is also lognormal.[3] This latter distribution is shown at the top of the figure and includes both the day-to-day and worker-to-worker components of variability in exposure. There is empirical evidence that these distributional assumptions are valid in many exposure situations, but the goodness of fit issue will be subsequently addressed in the context of the benzene data.

In order to apply this model to the analysis of the benzene data, it is necessary to give its analytical formulation and specify the parameters that will be estimated from the data. To this end, let $h(x)$ be the overall distribution of exposure levels (the distribution at the top of the figure), $f(x|\mu)$ be the distribution of the exposure of an individual worker whose mean exposure is μ (the distributions in the middle of the figure illustrating several values of μ), and $g(\mu)$ the distribution of means across the worker population (the distribution at the bottom of the figure). These three distributions are related by the expression:

$$h(x) = \int_0^\infty f(x|\mu)g(\mu)d\mu \qquad (1)$$

To explore the properties of $h(x)$ it is necessary to specify the parameters associated with *two* basic lognormal distributions. Specifically,

$g(\mu)$ is lognormal:
geometric mean = GMB
geometric standard deviation = GSDB

$f(x|\mu)$ is lognormal:
geometric mean = GMW
geometric standard deviation = GSDW

Further, represent the means and variances of the normal distributions of $\log(\mu)$ as $\mu_B = \log(GMB)$, $\sigma_B = \log(GSDB)$, and $\log(x)$ as $\mu_W = \log(GMW)$ and $\sigma_W = \log(GSDW)$. Note that B refers to the distribution between individuals which is the distribution at the bottom of Figure 1. Similarly, W refers to within individual distributions and T to the total distribution, i.e., the distributions in the middle and at the top of Figure 1, respectively.

Under these conditions (g and f both lognormal), h(x) is also a lognormal distribution with geometric mean equal to $\exp(\mu_B - \frac{1}{2}\sigma_W^2)$ and geometric standard deviation $\exp(\sqrt{\sigma_W^2 + \sigma_B^2})$. This is not a particularly well known result, but one that is to be found in texts on the subject.[3] The expected value of the exposure of a random individual selected from the distribution h(x) is $\exp(\mu_B + \frac{1}{2}\sigma_B^2)$, which is identical to the mean of the distribution $g(\mu)$. This can be seen in Table IV which contains an illustration of these three distributions for specific values of the lognormal parameters. The value of GSDT, for example, is found at the intersection of the **GSD** column and the total row and is equal to 3.666. It is the estimation and manipulation of these parameters that allows separate assessment of exposure due to individual differences (GSDB) versus those due to common environmental factors (GSDW).

TABLE IV
Example of the relationship between distributional parameters in the lognormal analysis of variance model

	Parameters				
Distribution	GM	GSD	μ	σ	Expectation
Between = $g(\mu)$	2.000	3.000	0.693	1.099	3.657
Within = $f(x\mid\mu)$	2.500	2.000	0.916	0.693	3.178
Total = h(x)	1.573	3.666	0.453	1.299	3.657

TABLE V
Empiric and parametric estimates for the proportion of samples exceeding 1.0 ppm

Group	P(X>1.0) (empiric)	P(X>1.0) (parametric)	Group	P(X>1.0) (empiric)	P(X>1.0) (parametric)
1	0.005	0.007	11	0.035	0.023
2	0.012	0.042	12	0.071	0.066
3	0.133	0.192	13'	0.063	0.045
4	0.138	0.154	14	0.112	0.090
5	0.091	0.046	15	0.124	0.123
6	0.167	0.122	16	0.022	0.012
7	0.200	0.184	17	0.186	0.193
8	0.255	0.260	18	0.250	0.274
9	0.121	0.102	19	0.145	0.112
10	0.113	0.108			

TABLE VI
Empiric and parametric estimates for the proportion of means exceeding 1.0 ppm

Group	$P(\mu>1.0)$ (empiric)	$P(\mu>1.0)$ (parametric)	Group	$P(\mu>1.0)$ (empiric)	$P(\mu>1.0)$ (parametric)
1	0.000	0.000	11	0.000	0.015
2	0.023	0.023	12	0.000	0.012
3	0.182	0.263	13	0.053	0.059
4	0.125	0.172	14	0.059	0.095
5	0.000	0.039	15	0.160	0.144
6	0.136	0.127	16	0.000	0.001
7	0.200	0.192	17	0.200	0.198
8	0.342	0.315	18	0.333	0.380
9	0.087	0.099	19	0.056	0.098
	0.120	0.106			

Highly sampled groups

In order to estimate the parameters of the model, it is clear that the exposure data have to include multiple measurements on each worker and multiple workers per group. In the benzene data, we required, somewhat arbitrarily, that each group analyzed should have at least 45 eight-hour measurements on at least **15** individuals with an average of at least ***two*** measurements per individual and not more than **30** percent of the measurements below the limit of detection. These criteria provide sufficient numbers of observations to produce stable estimates of the lognormal parameters.

Standard statistical procedures were used to estimate the model parameters.'') These procedures provide consistent estimates while taking account of the fraction of the measurements below the limits of detection. Before presenting these results, let us deal with the goodness of fit issue; that is, the degree to which the model adequately summarizes the data. The right tail of the top, h(x), and bottom, $g(\mu)$, distributions of Figure 1 were chosen as the focal points, clearly the most critical part of any exposure distribution. In particular, the method of choice was to contrast the proportion of measurements above 1.0 ppm predicted by the model with that observed and to contrast the predicted proportion of individual means about 1.0 ppm with those observed in the data set. That is, the means and variances of the model were estimated from the data, and these estimates were used to predict the fraction of measurements or the fraction of individual means above 1.0 ppm. These estimates were then contrasted with the observed exceedances. These contrasts are presented in Tables V and VI. These values are calculated using the relationships in the previous section. For example, the values in Table V come from h(x) where:

$$p = P\{Z > [\log(1.0) - \log(GMT)]/\log(GSDT)\}$$

where Z has a standard normal distribution.

As can be seen from Tables V and VI, the observed and estimated exceedances are generally in close agreement. Where there are differences, it is not clear whether they are due to inadequacies of the model or problems inherent in the data. That some data-related differences should exist is not surprising since, as noted earlier, these data were collected for many reasons, and it is unlikely that considerations of representativeness and independence were high priority criteria. In other words, some caution should be exercised in attempting to form useful conclusions based on statistical analysis of data that were not originally collected with such analyses in mind.

The benzene data are right-skewed and bounded below by zero, both caracteristics of the lognormal distribution. These facts, together with the data presented in these tables, are the principal evidence that can be offered on the distributional questions in any rigorous sense. The quantity of independent data was not sufficient to allow a more thorough evaluation of the distributional properties of the observed data. Nevertheless, the authors felt comfortable in adopting the lognormal model as a basis for subsequent analyses.

Table VII contains the results of an analysis of variance for the 19 groups. The mean value reported in Table VII is the estimated arithmetic mean of both the total distribution and of the distribution of individual means. This value and the appropriate GSD allow one to calculate the geometric mean for either the total distribution, $h(x)$, or the distribution of means, $g(\mu)$.

The percent column of Table VII gives the percent of the total variability that is attributable to differences between workers as opposed to that associated with the common environment. (This value is based on the variance of the logs of the measurements rather than the GSDs or the second moments of the lognormal distributions.) If this number is near 100 percent, it indicates that the group was heterogeneously exposed and that there were differences in the exposure of individual workers due to different work practices or to different tasks. A percent value of 100 corresponds to a GSDW of 1.00 which, in turn, corresponds to zero variability associated with the environment. As can be seen from Table VII, there are groups where the variability is principally environmental, e.g., Group 1, and others where variability is totally between workers. In general, however, a somewhat greater proportion of the total variability seems to be associated with differences between workers rather than with the common environment.

Three out of four cases in which the variability (GSDT) is almost totally associated with differences between workers are associated with groups from a single company. It was suspected that this might be due, at least in part, to the sampling strategy used by this company. For example, workers might have been preferentially selected for monitoring because their exposure was highly variable. To explore this issue, the correlation between the number of measurements per individual and the variance of those values for each group were calculated. Correlation coefficients are given in Table VII under the column labeled r. As can be seen, there is a clear tendency for high values of r to be associated with high values of GSDT, indicating that the total variability was inflated in some of these groups by repeated sampling of workers with the most variable exposures.

For some of these groups, this issue was pursued further and the possibility was explored that these high variances were associated with nonroutine exposure conditions. For Groups 5, 6, 7, and 8, all associated with one company, it was possible to determine that measurements taken during turnaround and inspection conditions were overrepresented in the sample and contributed disproportionately to the variance estimates. Similarly, the most extreme value of GSDT, 9.62, comes from Group 18 and results from the dominance of the data from only three individuals who have highly variable levels of exposure and were sampled repeatedly (r = 0.402). As mentioned previously, these results underscore the sensitivity of the analysis to the manner in which data are collected. If these data had resulted from a random and independent sampling process, one might conclude that these high variance groups were comprised of workers who had very different exposure experience and that control of exposures should be focused at the individual level. As it stands, these high variances seem to result largely from the way in which data were collected.

TABLE VII
Parameters of the lognormal model for benzene exposure and results from the one-way analysis applied to 19 specific job groups

Group	GSDB $g(\mu)$	GSDW $f(x\|\mu)$	GSM $h(x)$	Percent	μ_0^*	r
1	1.87	2.83	3.36	26.6	0.09	0.041
2	2.26	2.52	3.43	43.7	0.28	0.106
3	3.36	3.42	5.62	49.2	0.96	0.065
4	1.91	1.54	2.18	68.9	0.67	0.101
5	3.62	1.98	4.30	77.9	0.24	0.275
6	6.52	1.00	6.52	100.0	0.68	0.327
7	5.52	1.00	5.52	100.0	0.23	0.316
8	6.31	1.26	6.41	98.5	2.24	0.311
9	2.28	2.01	2.94	58.0	0.49	0.030
10	2.67	2.44	3.76	54.7	0.48	0.182
11	2.11	1.69	2.49	67.0	0.26	0.051
12	2.02	1.97	2.66	51.5	0.26	0.088
13	3.11	1.66	3.46	83.5	0.32	0.145
14	2.41	2.08	3.15	59.0	0.46	0.103
15	4.15	2.22	5.12	76.0	0.61	0.130
16	1.56	1.73	2.02	40.1	0.26	0.031
17	1.75	1.54	2.03	62.3	0.73	0.016
18	9.29	1.49	9.62	96.9	6.07	0.402
19	3.01	2.72	4.43	54.9	0.44	0.281

* This column is included so the entire parametric model may be reconstructed from the given data.

Conclusions

From a descriptive point of view, it was found that average eight-hour TWA exposures in most job groups that were analyzed were below 1.0 ppm. However, in a number of cases, the variability was sufficiently high to result in 1.0 ppm exceedance fractions of 0.1 or above. This finding is true for both the distribution of means, $g(\mu)$, and the distribution of measurements, $h(x)$, in those cases where adequate data were available to partition the variability. In a subsequent paper, the implications of this finding to health risk assessment and to compliance monitoring will be explored.

Although the data were not adequate to allow a conclusive investigation of the goodness of fit of the lognormal model, this limited exploration of the issue gave no indication that the assumption was misleading in this application. In fact, the conceptual clarification arising out of the analysis of variance model is a principal result of the benzene study. If the data have been collected so as to be representative of the exposure of the group, the lognormal model offers the hygienist an opportunity to identify the most promising strategy for exposure reduction. If, for example, it is found that there is little difference in the exposure of the workers in the group, then environmental controls are indicated. Conversely, large differences in the exposures of the individual workers suggests that an analysis of job tasks or of work practices may be more useful.

Recommendations

In characterizing the exposure of a group of workers to an airborne contaminant, sampling resources are most efficiently utilized if measurements are collected such that the tools of statistical inference can be used in the analysis of the data. This implies that exposure measurements should be made with due regard for randomness and independence. If, on the other hand, sampling strategies are used which focus on workers with particularly high or variable exposures, then analyzing such data as representative produces a distorted picture of the exposure of the group as a whole. This experience has shown that by assuming from the outset that each worker's exposure has both an individual and a common environmental component, one is led to an exposure model that appears to summarize the exposure data adequately and in a format that provides guidance in identifying strategies for diminishing exposures. While it cannot be recommended that this approach be applied to all exposure situations, it does provide a conceptual point of departure for a more detailed understanding of the factors underlying the variability in workplace exposures to toxic air contaminants.

References

1. Spear, R.C.; Selvin, S; Francis, M.: The Influence of Averaging Time on the Distribution of Exposures. Am. Ind. Hyg. Assoc. J. **47:365 (1986).**
2. Oldham, P.D.; Roach, S.A.: A Sampling Procedure for Measuring Industrial Dust Exposure. Br. J. Ind. Med. **9:112 (1952).**
3. Aitchison, **J.;** Brown, J.A.C.: The Lognormal Distribution, pp. **110–111.** Cambridge University Press, Cambridge **(1976).**

Received 10/22/86; review/decision 12/08/86; revision 2/10/87; approved 2/19/87

Answers to the "Action Level" questions

1. *c*
2. a
3. *c*
4. omission of water vapor, use of incorrect values for alveolar ventilation
5. *c*
6. a
7. "how safe" situations