STATISTICAL ANALYSIS OF BERYLLIUM EXPOSURE MONITORING RESULTS

An exposure monitoring and control strategy will depend in large part on the amount of variance in exposure levels. Analysis of the variance in exposures can help identify the important determinants to use in developing a monitoring and control strategy. In general, we find that exposures in even well-controlled beryllium activities will have more variance than is typical for occupational exposure data. Variance of beryllium exposure as measured by geometric standard deviation (GSD) has generally been greater than 3. This variance has primarily been within-worker variance or day-to-day variance, rather than between-worker variance. Within-worker variability would point to work practices as being an important determinant while between-worker variability would point to process equipment as being an important determinant.

Below are three example analyses of 8-Hr TWA beryllium exposure distributions. The first data set is made up of 529 breathing zone samples collected in a 1-month period in a Rocky Flats machine shop that fabricated beryllium metal parts. This shop had recently been associated with a high prevalence of CBD. It received a high level of industrial hygiene attention to both process controls and work practices leading to a more than 20-fold reduction in exposure levels. This is population data rather than a sample since each worker was monitored for each shift and there were no non-detected results. All distribution parameters are directly calculated. The geometric standard deviation for this group is 3.2. Notice that the arithmetic mean is about 3 times the geometric mean (and median). This demonstrates the relatively large influence excursions are having on the arithmetic mean.

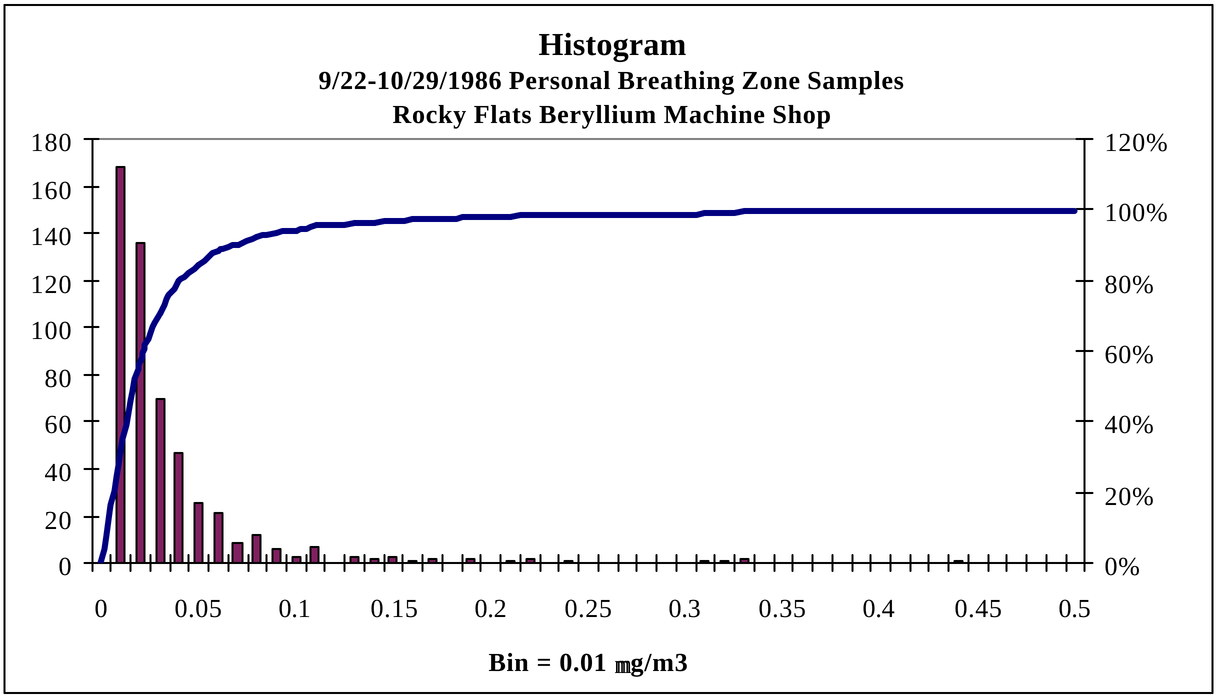
Descriptive Statistics
9/22-10/29/1986 Personal Breathing Zone Samples
Rocky Flats Beryllium Machine Shop

Geometric Mean	0.016 ì g/m^3
Geometric Standard Deviation	3.20
Arithmetic Mean	0.044 ì g/m^3
Actual 95th %	0.107 ì g/m^3

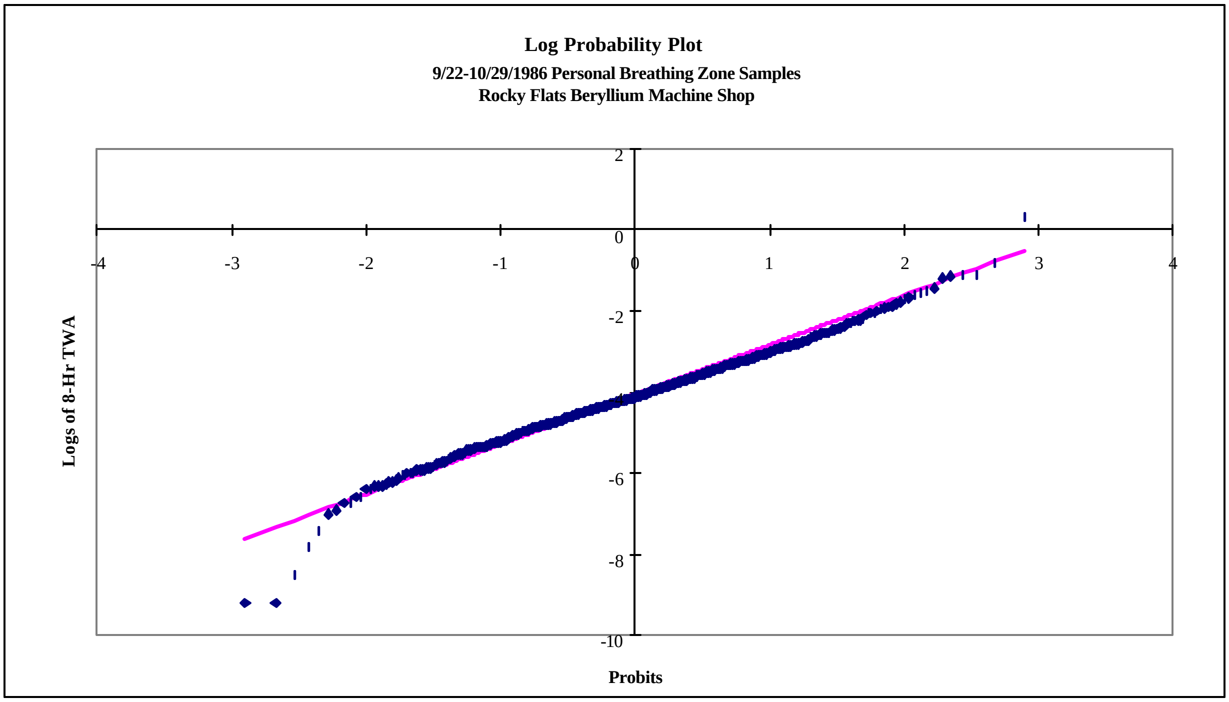
Range of 8-Hr TWAs

Rank		Result
1	-	0.0001 ì g/m^3
. . .		
265	-	0.016 ì g/m^3
. . .		
529	-	5.58 ì g/m^3

On a linear scale, you can see that the distribution of exposures is highly skewed.



The log-transformed data appear to be normally distributed, justifying the use of log normal statistics.



The group included 23 individuals who had between 12 and 28 measurements each. Below is an analysis of variance of log-transformed data performed by Microsoft Excel (Tools, Data Analysis, ANOVA.) This is the method the AIHA's "A Strategy for Assessing and Managing Occupational Exposures" (ref. 6 of this Guide) recommends for analyzing variance. Each "group" is the arbitrarily assigned ID number of an individual worker.

Anova: Single Factor

SUMMARY

Groups	**Count**	**Sum**	**Average**	**Variance**
10400	24	-79.9829	-3.33262	1.254854
12222	19	-68.6356	-3.6124	0.463363
12345	22	-106.445	-4.83841	0.514467
13333	20	-79.8691	-4.83841	0.514467
13456	17	-86.2928	-5.07604	1.312188
14444	24	-109.447	-4.56031	1.148093
14567	24	-76.1178	-3.17158	0.658827
15555	21	-84.6826	-4.0325	0.45387
15678	23	-121.068	-5.26382	1.263425
17890	24	-116.838	-4.86826	1.67401
18901	22	-74.5334	-3.38788	1.452094
19012	19	-84.1953	-4.43133	1.061457
22443	26	-105.691	-4.06503	1.393124
22451	23	-82.3002	-3.57827	0.588067
45491	24	-93.6838	-3.90349	1.161093
46979	20	-74.271	-3.71355	2.625731
50435	28	-129.372	-4.62042	0.326688
67709	28	-109.801	-3.92146	1.688668
76744	23	-102.826	-4.47069	1.544067
89177	28	-113.572	-4.05613	0.830275
95335	22	-85.5841	-3.93564	1.317352
99417	27	-93.7388	-3.47181	1.659504
516789	12	-43.327	-3.61058	0.39016

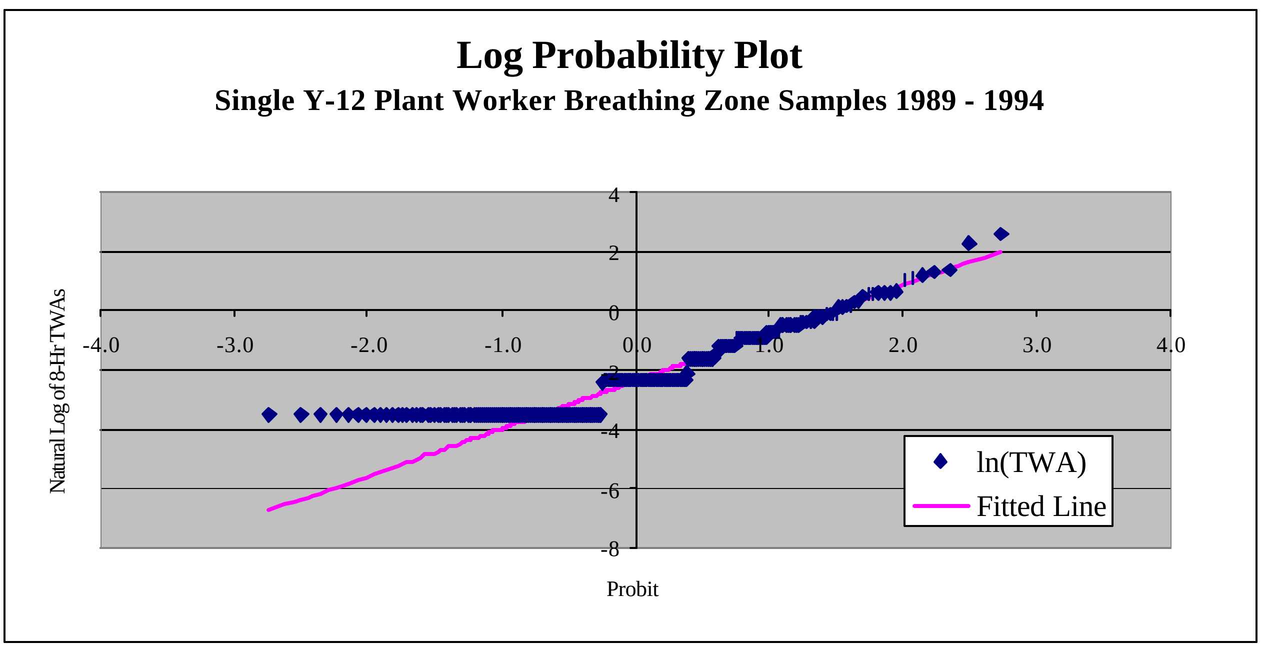
ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*Fcrit*
Between Groups	167.3707	22	7.607757	6.762879	8.17E-18	1.563627
Within Groups	559.0896	497	1.124929			

Two results are important. The F-Critical statistic is less than F, showing that individuals do not have the same mean exposure levels and are not a homogeneous exposure group. Exposure measurements from one worker are not representative of the exposures of other members of the group. The sum of squares statistic for within-worker variation is much larger than the same statistic for between-worker variation. Despite the fact that this was a production operation, work practices rather than process variables are the most important exposure determinant. This is probably due to the successful control of leakage from process equipment, which minimized process variables as a determinant of exposure levels.

This underlying distribution of exposures indicates the need for a high frequency of exposure monitoring. The large GSD and separation of the arithmetic mean from the median indicates that monitoring must be oriented toward detecting the infrequent excursions responsible for much of the health risk associated with this operation. Measurements from one worker are not representative of others. Any further reduction of exposures will depend primarily on ensuring that employees and their supervisors understand the work practices that are causing exposures. This requires monitoring data since we are operating in a realm where the senses are of little use in judging exposure potential.

A second example comes from the Oak Ridge Y-12 Plant. In this data set, not all shifts were monitored, and a high percentage of monitoring results were non-detect. A few individuals were monitored frequently, making it possible to estimate the degree of within-worker variance for them. Geometric mean and geometric standard deviation are estimated by using Microsoft Excel's regression function with the log transformed 8-Hr TWAs as the dependent variable and probits as the independent variable. Probits are the standard normal variable (z) calculated from the probability produced by dividing the rank order of the result by n+1. The Excel function NORMINV returns the probit that corresponds to the probability. The geometric mean is the exponent of the regression intercept and the geometric standard deviation is the exponent of the regression slope. The fitted line on the log probability plot is produced by multiplying the regression slope (X variable) times the probit plus the regression intercept. This computerizes the graphical method for estimating distribution parameters of censored data recommended in the AIHA's "A Strategy for Assessing and Managing Occupational Exposures" (ref. 6 of this Guide).



Distribution parameters are estimated based on the assumption that non-detected and non-measured 8-Hr TWAs would have also been log normal and fallen on the line fitted to the detected monitoring results. For this individual, 188 of 314 monitoring results were above the detection limit of 0.1 ì g/m^3. The regression output and estimated distribution parameters are as follows.

From the Excel regression function (Tools, Data Analysis, Regression)

Intercept	-2.38
X Variable	1.61

Estimated Distribution Parameters		Method of Calculating Estimate
Geometric Mean	0.09 ì g/m^3	By EXP of Regression Intercept
Geometric Standard Deviation	5.01	By EXP of Regression X Variable
Arithmetic Mean	0.34 ì g/m^3	By EXP(ln GM + ½(ln GSD)^2)
95th Percentile	1.32 ì g/m^3	By EXP(ln GM + 1.645(ln GSD))
Z value of 2 ì g/m^3	1.91	By Z = (ln 2 - ln GM)/ln GSD
Percent less than 2 ì g/m^3	97%	By Excel NORMSDIST(Z)
95/95 Geometric Upper Tolerance Limit	1.80 ì g/m^3	By EXP(ln GM + K (ln GSD)) Where K =1.84

Range of 8-Hr TWAs

Rank	Result
1	< 0.1 ì g/m^3 (Minimum)
. . .	
127	0.1 ì g/m^3 (First Detectable Result)
. . .	
157	0.1 ì g/m^3 (Median)
158	0.1 ì g/m^3 (Median)
. . .	
314	13.6ì g/m^3 (Maximum)

The Microsoft Excel ANOVA function cannot be used on highly censored data. If fewer than 10 percent of samples are below the detection limit, then substituting 2/3 the detection limit for the non-detected result has been recommended as an effective method of developing estimates, and this would allow the use of the ANOVA function. Alternate and fairly simple methods of analysis are discussed in Patty's Industrial Hygiene and Toxicology (see Rappaport, S.M. "Interpreting Levels of Exposures to Chemical Agents" in Patty's Industrial Hygiene and Toxicology, Vol. III, Part A, 3rd Edition; Harris, Cralley, and Cralley Eds, pages 349 - 404.) The GSD of the arithmetic mean exposure level of members of a group provides a measure of between-worker variability while the GSD of individual exposure levels is a measure of within-worker variability. A group is considered to be homogeneous if the 95 percent of individual mean exposure levels are within a factor of 4.

$$97.5\%\ \text{mean}/2.5\%\ \text{mean} < 4$$

Three other workers in the Y-12 Plant data set had large enough numbers of detected exposure monitoring results to estimate exposure parameters.

Rank	ln(mean)	Total Samples	Detectable Samples	GM	GSD	Mean
1*	-1.71	122	35	0.02	8.38	0.18
2	-1.08	314	188	0.09	5.01	0.34
3	-0.75	102	66	0.12	5.10	0.47
4	0.27	47	23	0.07	10.92	1.31

GSD of the Arithmetic Means 2.29

* *In his presentation at the 1999 AICHE "Exposure Estimation From Left-Censored Exposure Distributions," N. Esmen reported that the graphical method provided reasonable estimates when at least 30 percent of samples were detected. For the first worker 35/122 = 29 percent are detected.*

Again, within-worker variability is very large. The distance between the geometric and arithmetic means points to the large contribution of excursions to the overall risk of the group. For worker 2, one measurement, 13.6 ì g/m^3, accounted for more than a 12 percent of his or her mean exposure level. This underlying distribution of high day-to-day variability and low predictability in exposures indicates a need for frequent monitoring to provide workers and their supervisors with information on work practices that cause exposure.

A third example shows that frequent exposure monitoring can lead to an exceptional level of exposure control despite high variability. This data set is made-up of 7672 personal breathing zone measurements collected over a 2-year period from a crew cleaning beryllium-contaminated facilities and equipment at the Rocky Flats Environmental Technology Site. There was no operating process equipment contributing to exposure. The industrial hygienists recognized that work practices would be the primary exposure determinants. They established a 100 percent monitoring strategy and made sample analysis results available the next day. In this way they provided the information needed to understand the causes of exposures and helped develop a very high level of skill in this crew.

Like the Y-12 data, this distribution is left censored. All work shifts were monitored but over 70 percent (5560/7672) were non-detected. With the exception of the 95th percent, which was directly measured, distribution parameters are difficult to estimate with any confidence. The exposure potential is not trivial as evidenced by a few very high results. The arithmetic mean of this distribution is probably about 0.031 ì g/m^3. Substituting zero for non-detected results produced a mean of 0.0306 ì g/m^3 while substituting the detection limit for non-detected results produced a mean of 0.0313 ì g/m^3. This indicates that it is likely that the mean is larger than the 95th percent, which again points to risk being determined by a few very high exposures. In this situation, frequent monitoring is the only feasible method of detecting the exposures that create risk so that their causes can be determined and steps taken to reduce risk.

Range of 8-Hr TWAs	
Rank	Result
1	< 0.001 ì g/m^3 (Minimum)
3837	< 0.001 ì g/m^3 (Median)
5561 7290	0.001 ì g/m^3 (First Detectable Result) 0.02 ì g/m^3 (95th %)
7671 7672 7673	11.27 ì g/m^3 12.92 ì g/m^3 57 ì g/m^3 (Maximum)

INTENTIONALLY BLANK