CYANAMID

R.W. McCollum

TO: W. Fead NA
G. Koehler NA
B. Gaffney - Chicago
J. Caporossi NA

The attached article is for your information,

The effect of analytical error on the interpretation of blood lead data obtained from population studies is discussed. Statistical analysis of measured blood lead values reported in four representative population studies shows over 40 percent of the variability in blood lead measurements is due to analytical error. Because of this analytical error, a large percentage of the observed blood lead measurements which exceed a threshold limit are "false positives" in that the true blood lead values for these individuals do not exceed the threshold. The EPA has suggested that the distribution of blood lead values could be used as a guide for determining the portion of a population at risk. "At risk" is defined as having blood lead values above a certain threshold. If this is done, the effect of analytical error which causes "false positives" should be considered properly to obtain an accurate estimate of the population at risk.

Effect of analytical variability on measurements of population blood lead levels

JAMES M. LUCAS, Ph.D., Consultant
E.I. Du Pont de Nemours & Company, Inc., Wilmington, DE 19898

introduction

The Environmental Protection Agency has suggested that the distribution of blood lead values could be used as a biological guide for determining the portion of a population at risk as defined by blood lead levels above a certain threshold.[1] A similar guideline has been adopted by the Commission on European Communities.[2,3] In general, these biological guidelines, based on the cumulative frequency distribution of blood lead values in a population, enable the determination of the portion of a population which exceeds a specified threshold blood lead value. Although this approach takes into account the inherent variability of true blood lead values of individuals in a population, it can be misleading if it fails to consider fully the impact of analytical errors on the measured blood lead values.

The size of the analytical error in blood lead determinations that can be encountered in a highly experienced laboratory is shown in a recent publication.[4] In this study, a blood sample obtained from a single donor was divided into 35 separate samples. These samples were submitted along with other blood samples to the Kettering Laboratory over a period of nine months. The mean measured blood lead value was 19.13 μg/100 g with a standard deviation of 5.72. Values ranged from 12 to 42 μg lead/100 g of blood. If the extreme value of 42 μg/100 g is rejected as an outlier, the coefficient of variation is 21 percent. If it is included, as it would be in a test where only a single sample is made on an individual, the coefficient of variation is above 25 percent. The Kettering Laboratory is accredited by the American Industrial Hygiene Association, has long experience with blood lead analysis, and has performed well in many interlaboratory comparisons. Larger analytic errors have been reported in other studies[5] so errors such as these are not uncommon. While the presence of this analytical variability has been recognized,[6] its effect on the biological guideline blood lead distribution and on the proportion of individuals who will exceed a specified threshold has not been assessed. This paper uses a simple analysis of variance model to show the large effect that analytical errors can have on the blood lead distribution and on the proportion of individuals who will exceed commonly used thresholds.

In an effort to define the factors accounting for variation in blood lead levels in populations, analysis of blood lead data obtained from a number of different studies have been compared. For these studies, we have estimated the individual-to-individual variation and the analytical error variance components. These variance component estimates are then used in our calculations which estimate the percentage of individuals who will falsely or truly exceed selected threshold blood lead levels. Our calculations show that for typical populations the percentage of "false positives", which are measured values that exceed the threshold value because of analytical error, is large.

variations in measured blood lead values

The total variation in measured blood lead level for a population contains several components – true blood lead variation between individuals, within-person variability (over time) for an individual and a number of analytical error components which includes both sampling and measurement error. For most of this paper a simple separation of the total variability into two components, an analytical error component and the remainder (called the individual-to-individual component), will suffice. The effects of the individual-to-individual variation and of the analytical error are schematically shown in Figure 1.

In Figure 1-A, a hypothetical but reasonable distribution of blood lead results for a population is shown. As blood lead measurements are approximately lognormally distributed,[1,5] the drawings represent the distribution of the logarithm of blood lead values. As drawn, this distribution includes both the variation in the true individual-to-

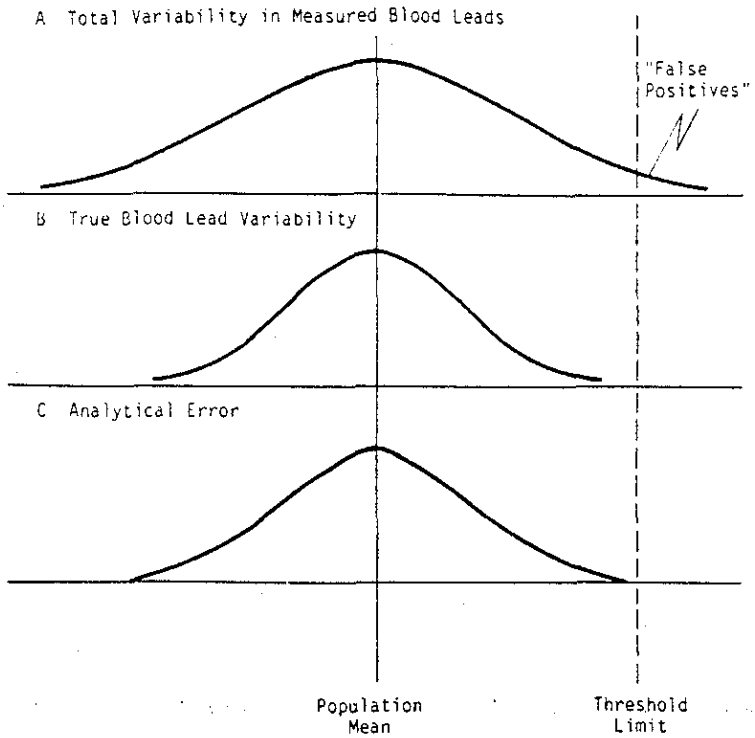

Figure 1 – Effects of variability on measured blood lead levels.

individual blood lead levels for the population group and the variation due to analytical error. In Figure 1-B, the hypothetical distribution which only includes true individual-to-individual variability in blood lead levels is pictured. In Figure 1-C, the hypothetical distribution which includes only analytical error is shown. This corresponds to the case where a single individual is sampled and analyzed repeatedly for lead. As shown in Figures 1-B and 1-C, almost none of the population is above the threshold, yet when the two effects are combined as pictured in Figure 1-A, "false positives" occur.

estimates of blood lead variability

The percentage contribution of the individual-to-individual variance and of the analytical variance to the total measured variance was determined for blood lead values from four major population studies. These studies were chosen as representative of the best recent studies, using venous blood samples, with repeat blood samples from individuals. A need for more research to determine all components of blood lead variability is indicated as there have been few studies in which repeat blood samples have been taken on individuals and no studies have been carried out to determine all components of individual-to-individual variability, and the components of analytical error. The four selected studies included a multi-city survey,[7] a longitudinal study,[8] a personal air monitor study,[9] and a Southern California rural-urban study.[10] Results of the analyses of variance on the studies are summarized in Table 1. The analysis of variance breakdown used to obtain Table 1 is described in detail in Appendix A. The analytical error was over 46 percent of the total measured blood lead variability in the four studies.

In the multi-city study, blood samples were obtained from more than 2000 suburban housewives in 10 communities across the United States. Duplicate samples were obtained from about 10% of the subjects. This study also measured air lead levels using stationary samplers, thus simulating how an air quality standard for lead would be implemented. In this study the value of the individual-to-individual variance was the smallest of all studies reported; it accounted for only

TABLE 1
Summary of Analysis of Variance for Blood Lead Values from Four Studies

	Component Standard Deviation	
	Individual-to-Individual	Analytical Error
Multi-City Survey[7]	0.052 (20%)*	0.103 (80%)
Longitudinal Study[8]	0.063 (12%)	0.174 (88%)
Personal Air Monitor Study[9]	0.097 (54%)	0.089 (46%)
Southern California Study[10]	0.087 (21%)	0.169 (79%)

*% Total Variance = $\left[\frac{\text{Component Standard Deviation}}{\text{Total Standard Deviation}}\right] \cdot 100$

All analyses are carried out on log blood lead levels

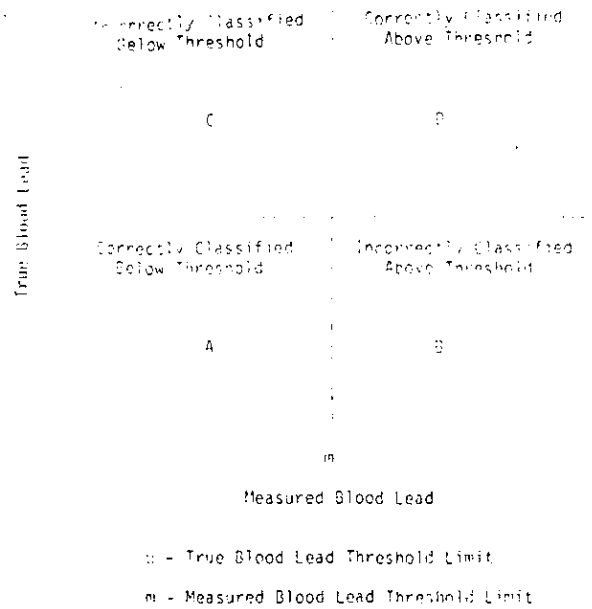

Figure 2 – True blood lead vs. measured blood lead.

20% of the total variance even through the method variance was not excessively large. For a homogeneous group of subjects, such as suburban housewives, the variability between individuals will be small.

The longitudinal study was a five-year study of more than 2000 workers from 23 locations around the United States. Urine samples were obtained from all subjects and blood lead samples were obtained from more than 200 individuals. On 217 individuals blood samples were obtained in each of the five years. No trends in blood lead were observed over the period of the study. Without duplicate blood samples in a year, long-term changes in an individual's blood lead cannot be separated from the analytical error. Thus, for this study the analytical error is inflated by long term within-person variability.

The personal air monitor study was conducted to determine the relationship between air lead levels and blood lead levels with the air lead being measured using personal air monitors. Five groups of thirty subjects each were monitored for a 2-4 week period. Two to eight blood samples were obtained from each subject and each sample was analyzed in duplicate. In three of the groups, two blood samples were taken each week of the study. With this sampling procedure week-to-week variability for an individual can be separated from analytical error. The analytical error component shown for this study does not contain this within-person variability. This study shows the smallest analytical error of all the studies. In fact, the analytical error estimated in this study is smaller than the measurement error shown in the Kettering Study[4] [where the measurement error is one of the components of analytical error as described in Appendix A]. To be conservative all our calculations showing the effect of analytical error use the variance components from this study.

In the Southern California study blood samples from two populations, an urban group and a rural group, were obtained. The urban group lived in a housing development close to a freeway (and a stationary air sampler which monitored air lead levels) while the rural group lived in an area of low lead exposure. While the study appeared to be well designed and conducted, there were often large differences in measured blood lead level between samples taken from the same individual even though these blood samples were taken only one day apart. For this study the analytical error was large.

estimation of "false positives"

When blood lead levels are measured on a typical human population, there are often individuals whose measured values are above a threshold value (e.g., 40 μg 100 mL). Repeat samples taken on these individuals usually have blood lead measurements at a much lower level – below the threshold value. This can occur because of analytical error or because of within-person variability (where an individual is above the threshold at the time of measurement but is below the threshold by the time of the follow-up sample). In this manuscript, we evaluate the effect of analytical error. Because of analytical error, a large percentage of the measurements that exceed a threshold are "false positives" in that the individual's true blood lead level is not above the threshold.

The occurrence of "false positives" is schematically illustrated by a bivariate plot shown in Figure 2 where the measured blood lead value is plotted on the X axis and the true blood lead value is plotted on the Y axis. The "measured threshold limit" and "true threshold limit" divide the distribution into four quadrants – "A", "B", "C", and "D".

"A" represents the measured values which are correctly classified as being below the threshold value since they lie below the measured and true threshold limits. The values in "D" are correctly classified as being above the threshold limits. The values in "B" and "C" are incorrectly classified. The values in "B" lie above the measured threshold but below the true threshold limit. The values in quadrant "B" will be incorrectly classified as being above the threshold: these are "false positives". In a similar manner, the values in "C" will be incorrectly classified as being below the threshold. They are "false negatives".

A numerical estimate of the percentage of "false positives" can be made by using a bivariate distribution model of blood lead values.[11] A summary of the results from such a calculation is given in Table II. The specific results in Table II were derived using the variance components from the personal air monitor study. The results from this study are conservative in that they show fewer "false positives" than the other studies because the data from this study had the smallest analytical errors. Similar results could be calculated for any study (or for other responses besides blood lead) when the individual-to-individual variance component and the analytical variance component can be determined. The description of the procedure used to obtain

TABLE II

Estimation of False and True Exceedences of Blood Lead Threshold Values

Population Geometric Mean (μg/100 mL)		15	20	25	30	35	40
Percent[a] of values observed above a threshold when the threshold is	30	1 11	9 06	27 38	50 00		
	35	0 26	3 25	13 36	30 56	50 00	
	40	0 06	1 11	6 06	17 14	32 98	50 00
Percent[b] of values which are "false positives" when the threshold is	30	1 05	6 77	12 76	11 80		
	35	0.26	2 86	8 93	13 07	11 80	
	40	0 06	1 05	4 91	10 41	13 19	11 80
Percent[c] of values actually above the threshold when the threshold is	30	0 10	3 50	20 76	50 00		
	35	0 01	0.62	6 63	24 54	50.00	
	40	0.00	0 10	1.78	9 92	27 53	50 00

[a]Percent in Categories B and D in Figure 2.

[b]Percent in Category B in Figure 2.

[c]Percent in Categories C and D in Figure 2.

the results in Table II using the logarithm of the blood lead values is given in Appendix B.

Table II shows that the analytical error will almost always cause some "false positives" when measuring blood lead levels in a population. When the population mean blood lead level is well below the threshold value, most of the values measured over the threshold will be falsely categorized. For example, in a population with a true mean blood lead level of 25 μg/100 mL (the upper end of the range for the average blood lead level in typical populations[11]), and a threshold value of 40 μg/100 mL, Table II shows that 6.06 percent of the individuals will be measured to be above the threshold. However, 4.91 percent of the population (81 percent of these individuals) will be "false positives" as their true blood lead is below 40 μg/100 mL.

The percentage of "false positives" increases as the population mean decreases for a given threshold value, although the absolute number of individuals classified above the threshold decreases. If a group has its mean near the threshold limit, an opposite problem occurs. In this group there can be many individuals whose measured blood lead is below the threshold limit while their true blood lead level is above the threshold. These would be "false negatives".

example of "false positives"

To illustrate how analytical error can account for the difference in the number of individuals above a threshold between the original sample and the follow-up sample, we will use the data from a recent survey in Pittsburgh that measured blood lead levels in children.[12] In this study a representative sample of households from the metropolitan Pittsburgh area was surveyed for levels of lead in paint and dust in homes. Following this survey, blood samples were requested from children in households having children aged 7 or younger. While not all households complied with the request, the complying households had higher lead content in dust and paint than the noncomplying households. Four hundred fifty-six children were sampled. The average of the 456 blood lead values was 21.9 (median 21.0) and the standard deviation was 7.9. Fifteen children (3%) showed blood lead values above 40 on the first sample. A resample was requested for these children. Fourteen of these 15 children were found for a resample. Most children were

TABLE III

Blood Lead Values for 15 Out of 456 Children in Pittsburgh Study Whose Initial Blood Lead Values Were Greater Than 40 μg/100 mL

Child	Measured Blood Pb (μg/100 mL) First Sample	Second Sample	Third Sample	Predicted Blood Pb (μg/100 mL)	
1	45	26		32	31
2[a]	73	73	75	41	39
3	42	35		31	30
4	43	27		31	30
5	45	38		32	31
6	41	28		30	29
7	46	34		32	31
8	42	38		31	30
9	47	35		33	31
10	59	28		37	35
11	42	19		31	30
12[a]	43	Could not be located		31	30
13	40	21		30	29
14[a]	50	48	43	34	32
15	41	32		30	29
RMS[b]	16 1			6 3	6 1

[b]First column – Regression caused by analytical variability. Second column – Regression caused by analytical variability and within-person variability.

[a]Not used in RMS calculation.

[b]RMS Root Mean Square Difference from second sample

resampled within one month of the original sample. There was no attempt to change exposures between the original sample and the resample. In the resample, 2 of the 15 children remained above 40 while 12 dropped below. These results are shown in Table III. The very high value was obtained from an eight-year-old being treated for lead poisoning.

Using the regression equation shown in Appendix C, we can predict the blood lead levels for individuals in the Pittsburgh study whose initial blood lead was above 40. The regression has two causes; the first is analytical error, and the second is within-person variability (as the blood lead level could be above the threshold at the time of the original measurement but below the threshold by the time a resample is taken). In Table III, we show the results of two regression calculations. One considers only the regression effect due to analytical error while the second gets slightly better predictions as it predicts the combined regression caused by both effects. The blood lead values predicted using the regression equations are much closer to the resampled blood lead values than are the initial sample values. The good agreement between the predicted and the resample blood lead levels shows the validity of the calculation of "false positive" percentages (Table II).

In summary, we have shown that the percentage of blood lead values measured to be above a threshold will usually overestimate the percentage actually above the threshold. For typical populations, the percentage of "false positives" will be substantial. A numerical estimate of the extent of "false positives" has been given along with a predictor that is a better estimate of a confirmatory blood lead reading than is the original reading.

The existence of false positives depends only on the fact that the total observed variability is bigger than variability in the true values due to the presence of analytical error. The Analysis of Variance model used here is more general than the classification models used elsewhere[13] as it shows how the classification probabilities change with changes in population mean level and with the threshold level. Thus, the approach illustrated here is appropriate for any quantitative variable.

acknowledgement

I would like to thank R.D. Snee for providing the original data for the personal air monitor study, for performing some of the analysis of variances described and for many helpful discussions.

references

1. **Office of Research and Development:** *Air Quality Criteria for Lead,* EPA-600/8-77-017. U S Environmental Protection Agency, Washington, D C (1977)
2. **Zielhuis, R.L.:** Biological Quality Guide for Inorganic Lead *Int. Arch. Arbeitsmed. 32* 103-127 (1974).
3. **Commission of the European Communities:** Council Directive of March 19, 1977, on Biological Screening of the Population for Lead. *Official Journal of the European Communities, Vol. 20, No L105* 10-17, Brussels (1977)
4. **Lerner, S.:** Blood Lead Analysis — Precision and Stability. *J. Occup. Med. 17* 153-154 (1975).
5. **Billick, I.H., A.S. Curran and D.R. Shier:** Analysis of Pediatric Blood Lead Levels in New York City for 1970-1976 *Environ. Health Perspect. 31* 183-190 (1979).
6. **Pierce, J.O., S.R. Koirtyohann, T.E. Clevenber and F.E. Lichte:** *The Determination of Lead in Blood.* International Lead Zinc Research Organization, Inc., New York (1976).
7. **Tepper, L.B. and L.S. Levin:** A Survey of Air and Population Lead Levels in Selected American Communities *Environmental Quality and Safety, Supplement Vol. II — Lead,* pp. 152-196. George Thieme Verlag, Stuttgart, W. Germany (1975).
8. **McLaughlin, M., A.L. Linch and R.D. Snee:** Longitudinal Studies of Lead Levels in a U.S. Population *Arch. Environ. Health 26* 131-136 (1973).
9. **Azar, A., K. Habibi and R.D. Snee:** An Epidemiologic Approach to Community Air Pb Exposure Using Personal Air Samplers. *Environmental Quality and Safety, Supplement Vol. II — Lead,* pp. 254-270. George Thieme Verlag, Stuttgart, W. Germany (1975).
10. **Johnson, D.E., R.J. Prevost and J.B. Tillery:** *Baseline Levels of Platinum and Palladium in Human Tissues.* EPA-600/1-76-019. Environmental Health Effects Research Science Document, U.S. Environmental Protection Agency, Washington, D.C. (1976).
11. **National Bureau of Standards:** *Tables of the Bivariate Normal Distribution Function and Related Functions.* Applied Mathematics Series No. 50, U.S. Government Printing Office, Washington, D.C. (1959).
12. **Urban, W.D.:** *Statistical Analyses of Blood Lead Levels of Children Surveyed in Pittsburgh, PA: Analytical Methodology and Summary Results.* NBSIR 76-1024. Institute of Applied Technology, National Bureau of Standards, Washington, D.C. (1976).
13. **Colton, T.:** *Statistics in Medicine.* Little Brown and Co., Boston, Mass. (1974).
14. **Scheffe, H.:** *The Analysis of Variance.* John Wiley and Sons, New York (1959).
15. **Snee, R.D. and P.E. Smith, Jr.:** Statistical Analysis of Interlaboratory Studies. *Am. Ind. Hyg. Assoc. J. 33* 784-790 (1972).
16. **Koppler, J.F., M.E. Maxfield, W.E. Moss, G. Tietjen and A.L. Linch:** Interlaboratory Evaluation of the Reliability of Blood Pb Analyses. *Am. Ind. Hyg. Assoc. J. 31* 412 429 (1970).

APPENDIX A

Estimates of the total variance and the components of the total variance for blood lead values were obtained for four studies using a standard hierarchical analysis of variance.[14] As the distribution of blood lead value is approximately log normal,[15,5] all results are developed on a $\log_{10}$ scale. The variance components for these studies are discussed in this appendix.

Consider a study where:

g = number of groups

TABLE A-1
Analysis of Variance

Source	Number	Degrees of Freedom	Variance Component
Groups	g	g - 1	
Person-to-Person	n	g(n - 1)	
Within-Person	t	gn(t - 1)	
Sampling	s	gnt(s - 1)	
Measurement	m	gnts(m - 1)	

$$\sigma^2_{analytical} = \sigma^2_{sampling} + \frac{\sigma^2_{measurement}}{m}$$

n = number of individuals (per group)
t = number of sampling times (per individual)
r = number of samples (per sampling time)
d = number of determinations (per sample)
σ^2 = variance component

The data from such a study could be analyzed using the analysis of variance, shown in Table A-1. Even more extensive breakdowns could be considered. For example, the measurement variance component could be further broken down by considering the component parts of the measurement error, such as the operator-to-operator error and the instrument-to-instrument error. Data to make this more comprehensive breakdown is seldom available.

In the analysis of variances shown in Table A-1, the analytical error contains two components, a sampling component and a measurement component as shown by the following formula:

$$\sigma^2_{analytical} = \sigma^2_{sampling} + \sigma^2_{measurement}/m$$

In most screening studies only a single measurement is made on each sample (m=1). When this occurs, the sampling variance component and the measurement variance component cannot be estimated separately; only the total analytical error can be estimated. Our estimate of the effect of analytical error will assume that a single measurement is made on each sample, so for our calculations

$$\sigma^2_{analytical} = \sigma^2_{sampling} + \sigma^2_{measurement}$$

Generally, enough samples are not taken to estimate all the variance components shown in Table A-1. For example, in a study in which each individual is sampled on two occasions and each sample is measured once, only two variance components can be estimated. These two components are:

1) $\sigma^2_{person} + (k)\sigma^2_{within\text{-}person}$

2) the sum of $(1-k)\sigma^2_{within\text{-}person} + \sigma^2_{sampling} + \sigma^2_{measurement}$

In this formula, the k value which multiplies the within-person variance component is a number between 0 and 1. If the two samples from an individual are taken close together, k will approach one; however, if there is a long time between samples then k will approach zero.

In this latter case, if the second component is interpreted as analytical error, it will be inflated as it contains much of the within-person variability.

When only a single sample is taken from an individual, but duplicate determinations are made on the blood sample, $\sigma^2_{measurement}$ can be estimated. However, the total analytical error cannot be estimated as no estimate of sampling variability can be obtained. In this case the measurement error may be (falsely) reported as the total analytical error. This causes the precision of the study to be overestimated. To estimate the analytical variability, more than one sample must be taken from an individual. Table A-2 summarizes analyses of variance for blood lead values from four studies for which multiple samples were taken from individuals. In the longitudinal study[8], the repeat samples from an individual were taken a year apart. For this study, the within-person variability is combined with the analytical error causing an overestimate of the analytical variability.

TABLE A-2
Analysis of Variance for Blood Lead Values from Four Studies

Study	Component Standard Deviation[1]			Geometric Standard Deviation
	Individual-to-Individual	Analytical Error	Total	
Multi-City Survey[7]	0.0521 (20)[2]	0.1030 (80)	0.1154	1.30
Longitudinal Study[8]	0.063 (12)	0.174 (88)	0.185	1.53
Personal Air Monitor Study[9]	0.0972 (67)	0.0688 (33)	0.1191	1.32
	0.0972[3] (54)	0.0889 (46)	0.1317	1.35
Southern California Study[10]	0.0871 (21)	0.1689 (79)	0.1900	1.55
	0.0784[4] (18)	0.1662 (82)	0.1838	1.53

[1] Component Standard Deviation = (Variance Component)$^{1/2}$

[2] Values in parentheses are % Total Variance = $\left[\frac{\text{Component Standard Deviation}}{\text{Total Standard Deviation}}\right]^2 \times 100$

[3] Modified $\sigma^2_{analytical}$ obtained from duplicates (see Equation (1))

[4] Deleted six outlier blood lead values.

TABLE A-3
Personal Air Monitor Study Analysis of Variance Table

Source	Degrees of Freedom	Variance Component	Variance Component Estimates
Group	4	0.00893	
Person	145	0.00854	
Sampling Time	195	0.00090	
Analytical	194	0.00474	$\sigma^2_{sampling} + \sigma^2_{measurement}/2$

When single readings are made

$$\sigma^2_{within\text{-}person} = \sigma^2_{sampling} + \sigma^2_{measurement} = (0.0889)^2$$

$$\sigma^2_{total} = \sigma^2_{person} + \sigma^2_{within\text{-}person} = 0.00944 = (0.0972)^2$$

$$\sigma^2_{total} = \sigma^2_{measurement} + \sigma^2_{analytical} \quad (0.0972)^2 + (0.0889)^2 = (0.1317)^2$$

For the other three studies described, duplicate samples were taken much closer together (so k = 1). For example, in the Southern California study[10] which had the largest percentage analytical error, duplicate samples were taken one day apart.

In the four studies summarized in Table A-2, a component estimating group-to-group differences was also obtained as each of the studies was made on more than one group. This component which accounts for differences between groups was not reported as part of the individual-to-individual variance.

In the personal air monitor studies, two blood samples were taken from most individuals during each week of the study. Thus, for this study it was possible to separate the week-to-week within-person variability from the analytical error. Table A-3 shows the detailed analysis of variance for the personal air monitor study.

In the personal air monitor study, duplicate determinations were carried out on each blood sample. Only the average of duplicate determinations was used in the analysis since the values of the individual determinations were not available. In the analysis, therefore, only the sum of $\sigma^2_{sampling}$ and $\sigma^2_{measurement}/2$ could be estimated. The following equation was used to estimate the analytical error which would be obtained if only single determinations were used:

$$\sigma^2_{analytical\ (single\ determination)} = \sigma^2_{analytical\ (duplicate\ determination)} + \sigma^2_{measurement} \cdot 2 \quad (1)$$

For this study:

$$\sigma^2_{analytical\ (single\ determination)} = 0.00474 + (0.0795)^2 \cdot 2 = (0.0889)^2$$

The value of $(0.0795)^2$ for $\sigma^2_{measurement}$ was obtained from an AIHA study.[15,16] This modification was used to make the personal air monitor study consistent with the other studies which used only single determinations.

The analytical variability for the personal air-monitoring study is the smallest of all four studies considered. In fact, the analytical error for the personal air-monitor study is smaller than the sampling error found in the Kettering Lab study.[3] Had the estimate of false positives used the variance components from any of the other studies, the percentage of false positives would have been greater.

APPENDIX B

calculation of bivariate lognormal distribution probabilities

The relationship between measured blood lead and true blood lead values can be described by a bivariate lognormal distribution (a bivariate normal distribution using log blood lead values).

Five parameters are required to describe this distribution:

- u_t = the mean of the true blood lead values
- u_m = the mean of the measured blood lead values
- σ^2_t = the true variance in blood lead values
- σ^2_m = the variance in measured blood lead values
- ρ = the correlation between measured and true blood lead values

The means of the true and the measured blood lead values are considered to be the same; they are equal to the population mean. A bias in the measurement could easily be accounted for by adjusting the mean levels. The true blood lead variance σ^2_t is estimated by $\sigma^2_{person} + \sigma^2_{within\text{-}person}$ which for the personal air monitor study is $(0.0972)^2$. The variance in measured blood lead levels (σ^2_m) is estimated by σ^2_{total} which is $(0.1317)^2$. The correlation (ρ) is:

$$\rho = \frac{\text{Covariance (XY)}}{[\text{Variance (X) Variance (Y)}]^{1/2}}$$

$$\rho = \frac{\sigma^2_t}{[\sigma^2_m \times \sigma^2_t]^{1/2}} = \frac{\sigma_t}{\sigma_m}$$

Thus our estimate of ρ is

$$\rho = \frac{0.0972}{0.1317} = 0.738$$

The probabilities in Tables II and B-1 were obtained by first calculating Z_1 and Z_2 where:

$$Z_1 = \frac{\log (\text{true threshold value}) - \log u_t}{\sigma_t}$$

$$Z_2 = \frac{\log (\text{measured threshold value}) - \log u_m}{\sigma_m}$$

Note that the measured threshold value and the true threshold value need not be equal.

Values of Z_1 and Z_2 were used with normal probability tables to determine the probabilities in quadrants A + C, B + D, A + B, and C + D. (These refer to quadrants shown in Figure 2.) The probability in quadrant D was obtained from tables of the bivariate normal distribution[11] using Z_1, Z_2

TABLE B-1

Estimation of False and True Exceedences of Blood Lead Threshold Values

Population Geometric Mean Blood Lead (μg/100 mL)	Threshold for Blood Lead (μg/100 mL)		Bivariate Distribution* (percent)				% Values Observed Above Measured Threshold, (B + D)	% Positives Which Are False Positives, $\frac{100\,B}{(B+D)}$	% True Exceedences of Threshold, (C + D)
	Meas	True	A	B**	C	D			
15	30	30	98 85	1 05	0 04	0 06	1 11	95	0 10
20			89 73	6 77	1 21	2 29	9 06	75	3 50
25			66 49	12 76	6 13	14 62	27 38	47	20 76
30			38 20	11 80	11 80	38 20	50 00	24	50 00
15	30	35	98 89	1 11	0 00	0 01	1 11	100	0 01
20			90 86	8 53	0 09	0 53	9 06	94	0 62
25			71 79	21 59	0 83	5 79	27 38	79	6 62
30			47 42	28 04	2 58	21 96	50 00	56	24 53
35			26 43	23 58	4 13	45 87	69 44	34	50 00
15	35	35	99 74	0 26	0 00	0 00	0 26	100	0 01
20			96 54	2 86	0 23	0 39	3 25	88	0 62
25			84 44	8 93	2 21	4 42	13 36	67	6 63
30			62 39	13 07	7 05	17 49	30 56	43	24 54
35			38 20	11 80	11 80	38 20	50 00	24	50 00
15	30	40	98 89	1 11	0 00	0 00	1 11	100	0 00
20			90 94	8 96	0 00	0 09	9 06	99	0 10
25			72 54	25 68	0 08	1 71	27 38	94	1 78
30			49 63	40 44	0 37	9 56	50 00	81	9 92
35			29 72	42 74	0 84	26 70	69 44	62	27 54
40			15 94	34 04	1 19	48 82	82 86	41	50 00
15	35	40	99 74	0 26	0 00	0 00	0 26	100	0 00
20			96 74	3 17	0 02	0 08	3 25	98	0 10
25			86 36	11 86	0 28	1 50	13 36	89	1 78
30			68 04	22 04	1 40	8 52	30 56	72	9 92
35			46 70	25 77	3 30	24 23	50 00	52	27 52
40			28 12	21 89	4 86	45 13	67 02	33	50 00
15	40	40	99 94	0 06	0 00	0 00	0 06	100	0 00
20			98 85	1 05	0 04	0 06	1 11	95	0 10
25			93 31	4 91	0 64	1 15	6 06	81	1 78
30			79 67	10 41	3 20	6 73	17 14	61	9 92
35			59 27	13 19	7 74	19 79	32 98	40	27 53
40			38 20	11 80	11 80	38 20	50 00	24	50 00

*A - Values correctly classified as below threshold.

B - Values incorrectly classified as above threshold.

C - Values incorrectly classified as below threshold.

D - Values correctly classified as above threshold.

** - % False Positives for the Measured Threshold.

and ρ. (Actually, a computer program that followed the above procedure was used, as interpolation would be required in the tables.) Once the probability in quadrant D is known, the probability in all quadrants can be calculated.

In Table B-1, the four classification probabilities of the bivariate distribution are presented for different population mean and assumed threshold blood lead values. In addition, the last three columns present the percent of values observed above a threshold, (B + D), the percent positives which are "false positives", (B/(B + D) × 100), and the percent exceedences of the threshold, (C + D).

APPENDIX C

prediction of blood lead values for repeat blood samples

The value of a follow-up blood lead sample can be predicted using the following regression equation which estimates the statistical phenomena of "regression toward the mean". Individuals whose measured blood lead levels are above a threshold will either be those who have positive analytical errors or those who have high true blood lead levels at the time measurements are taken. When only these individuals are resampled, the resample will tend to be lower because

positive analytical errors will occur less often or because the individual's blood lead may have dropped in the time between the original sample and the resample. Using the estimated variance components from the personal air monitor study we will estimate the regression effect caused by analytical error and the regression effect caused by the combined effect of analytical error and within-person variability. The regression calculation enables one to make a good prediction of the resample blood lead value.

The regression equation has the form

$$\log Y = \log u + \beta(\log X - \log u) \qquad (1)$$

where:

- Y = the predicted blood lead value for the repeat blood sample
- X = the original blood lead reading
- u = the geometric mean of the population (assumed to be 21 for the Pittsburgh study)
- β = regression coefficient

For the personal air monitor study the regression coefficient is:

$$\beta = \frac{\sigma^2_{\text{person}}}{\sigma^2_{\text{total}}} = \frac{0.00854}{0.01734} = 0.493$$

for the regression equation that includes both the effect of analytical error and the effect of within-person variability and

$$\beta = \frac{\sigma^2_{\text{person}} + \sigma^2_{\text{within-person}}}{\sigma^2_{\text{total}}} = \frac{0.00944}{0.01734} = 0.544$$

for the regression equation that includes only the effect of analytical error

- σ^2_{total} = the total variability (for single samples) = $(0.1317)^2$
- σ^2_{person} = person-to-person variability = 0.00854
- $\sigma^2_{\text{within-person}}$ = the within-person variability = 0.00090

The variance component values from the personal air monitor study that were used as the analytical error could not be estimated from the Pittsburgh data as there were no duplicate samples.

CYANAMID

INTEROFFICE CORRESPONDENCE

Chicago Plant (OFFICE) (BLDG TUBE RM#) 10/2/79 (DATE)

TO: Bound Brook

ATT'N. OF: M. G. Duffy

COPY TO: G. J. Brucia - NA
E. J. Dean - NA
F. B. Dorf - NA
H. T. Walker - NA

SUBJECT: Blood Lead Incentive

REFERENCE:

Presently at the Chicago Plant the employees receive one book of S&H Green Stamps ($3.00 value) when their latest blood lead analysis is below 0.06 mg lead/100 grams of blood. Based on the monitoring time table the maximum award/employee is six books of stamps/year. However, there is also a stipulation that if all employees are below 0.06 all would receive two books of stamps - This situation has occurred only once.

It is proposed that this incentive system be replaced by one of the following or a combination. This would be based on an employee achieving blood lead levels below 0.05 mg lead/100 grams of blood.

A. An employee would receive one day off with pay taken within one month of the test and in agreement with plant management.

B. An employee would receive one days pay with no time off.

C. An employee could choose a gift from a catalog. Prizes in the range of a days pay.

D. An employee could chose one of the above.

H. C. Gaffney

HCG:ck

NA 1128 REV. 9-69 9-69

CYWI 3-000953

N14729.02