UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
REGION VIII (8HWM-SM)
999 18th STREET - SUITE 500
DENVER, COLORADO 80202-2466

Region VIII

REGION 8 SUPERFUND TECHNICAL GUIDANCE

No. RA-06: Exposure to Indoor Dust Sep 1995
Risk Assessment (Short Title / Key Words)

TITLE: *Estimating Site-Specific Exposure to Contaminants in Indoor Dust*

authors: Region 8 Toxicologists
contact: Dr. Susan Griffin
ph: 303-294-7655

approvals: (initials)
PSA Chief, TS /date
CLC / 9/28/95 Chief, SM /date
TH for Dupray 9.29.95 Director, HWM /date

HAZARDOUS WASTE MANAGEMENT DIVISION, SUPERFUND MANAGEMENT BRANCH, TECHNICAL SECTION

SUMMARY

Exposure to contaminants in indoor dust is a significant exposure pathway which is often overlooked in a human health risk assessment. This Regional Guidance provides a basic mathematical approach for estimating site-specific exposure from contaminants in indoor dust for use in the baseline risk assessment.

EVALUATING EXPOSURE FROM INDOOR DUST

OBJECTIVE

The objective of this Regional Guidance is to provide guidance on estimating site-specific exposure to contaminants in indoor dust for the human health baseline risk assessment. This guidance reviews the evidence for exposure to indoor dust, and provides information on how and when to deviate from standard default assumptions on dust exposure.

BACKGROUND

Evidence That Dust Exposure is Significant

Direct evidence that dust exposure is a pathway of potential concern comes mainly from blood lead studies in which it is possible to investigate the relationship between the level of lead in dust and the concentration of lead in blood. Typically, paired data (i.e., data collected from the same location) are collected for blood lead and dust lead, and these data are fit to a simple bi-variate model of the form:

$$PbB = PbB_0 + k_1 \cdot C_d$$

where:

PbB = Observed blood lead concentration (ug/dL) in a person living at a particular location

PbB_0 = Concentration of lead in blood (ug/dL) in that person not attributable to dust ingestion

k_1 = Mean increase in blood lead per unit concentration in dust lead (ug/dL per ppm)

C_d = Concentration of lead in dust of the house where the person lives (ppm)

Example results are presented below in Table 1.

Even stronger correlations are observed when hand-loading (ug/cm^2) rather than

Table 1

Site	k_1 (ug/dL per 1000 ppm)	Reference
Kellogg	1.36	Marcus and Bernholic 1990
East Helena	1.26	Marcus and Bernholic 1990
Leadville	1.4	Weston 1995
Rochester	0.14	Lanphear et al. 1994

dust concentration is used (Lanphear et al. 1994), and advanced structural modeling confirms the importance of dust exposure, especially when the influence of behavioral and demographic variables (e.g., mouthing frequency, age, socioeconomic status) are accounted for (Bornschein 1985, Marcus and Bernholic 1990).

Default Assumptions for Residential Exposure to Dust

Contamination of dust can lead to human exposure via a number of pathways, including ingestion, dermal contact, and inhalation of particles and/or vapors. Of these pathways, incidental ingestion is usually considered to be the most significant. Current USEPA national guidance on ingestion exposure of residents to dust specifies the following default assumptions:

1) Upper bound (RME) estimates of the total amount of soil plus dust ingested each day are 200 mg/day by a child, and 100 mg/day by an adult (EPA 1991, 1993). Central tendency estimates are 100 mg/day

by a child and 50 mg/day by an adult (EPA 1993). For a child, it is assumed that 55% comes from dust and 45% comes from soil (EPA 1994).

2) For risk calculations, the concentration of contaminants in dust should be assumed to be equal to that in soil ($C_d = C_s$) (EPA 1989). For calculations of exposure and risk to lead contamination, the recommended default assumption is $C_d = 0.7*C_s$ (EPA 1994).

However, data obtained at several Region VIII mining/smelting sites indicates that the concentration of metals (e.g., lead, arsenic) in indoor dust is often less than either the RME or the central tendency default assumptions would predict. This Regional Guidance reviews the studies which support this conclusion, and provides guidance on the type and extent of data that are needed to revise default assumption 2.

DISCUSSION

Conceptual Model (Basic Mathematical Approach)

Contaminants can presumably enter a house and contribute to indoor dust by two main pathways:

- Penetration of airborne particles through open doors and windows, followed by deposition onto indoor surfaces

- Physical transport of soil indoors, mainly from material adhering to shoes, clothing, house pets, etc.

In both cases, the source of the contamination can be partly from the immediate vicinity of the house, and partly from areas outside the yard boundary.

The amount of contamination entering a house via these mechanisms presumably depends on a large number of independent variables such as frequency of entering and leaving the house, presence or absence of a "mud room", tendency to have dirt on shoes or clothing, presence or absence of pets, the amount of soil particles in ambient air, wind direction from area sources, temperature patterns (which influence the tendency to have doors and windows open), the typical size of contaminated soil particles, etc.. In addition, metal contamination can occur in indoor dust as a consequence of release from non-soil sources. For example, lead in indoor dust can be derived from indoor leaded paint. Thus, it is expected that the contribution of soil to indoor dust may vary widely, both between different sites and at different homes within a site.

Mathematical Model for Data Evaluation

Based on the assumption that dust contamination can arise from three types of source (area sources, local yard soils, indoor sources), the following mass-balance equation describes the concentration of contaminant in indoor dust:

$$C_d = \frac{(m_s \cdot C_s + m_a \cdot C_a + m_i \cdot C_i)}{(m_s + m_a + m_i)} \quad (1)$$

where:

m = mass of material in dust derived from yard soils (m_s), area sources (m_a), or indoor sources (m_i).

C = concentration of contaminant in interior dust (C_d), yard soils (C_s), area sources (C_a), or indoor sources (C_i).

This equation can be re-written as follows:

$$C_d = k_a C_a + k_s C_s + k_i C_i \quad (2)$$

where:

k = mass fraction in dust of material derived from area sources (k_a), yard soil (k_s) and indoor sources (k_i).

In theory, this basic equation can then be used to estimate the best fit values of k_a, k_s, and k_i from a site-specific data set in which paired data are available for the concentration of contaminants in dust, yard soil and indoor sources (e.g., paint). In practice, however, data are rarely available on the concentration of chemical in indoor sources, although related measures are sometimes available (e.g., the amount of lead in paint is often quantified in units of mg/cm^2). In addition, the contribution of indoor sources such as paint depends not only on the concentration, but also on the age and condition of the paint. Consequently, it is frequently difficult to separate the relative contribution of indoor sources from those of area sources. Thus, it is usually simplest to analyze paired soil-dust data using a simple two-parameter model, as follows:

$$C_d = k_0 + k_s \cdot C_s \qquad (3)$$

where:

k_0 = contribution to indoor dust from non-yard soil sources (ppm)
k_s = mass fraction of yard soil in indoor dust (unitless)
C_s = Concentration in yard soil (ppm)

Effect of Measurement Error

The most appropriate way to analyze and quantify the soil/dust relationship depends on whether or not there is significant measurement error in the data. Measurement error is the difference between the true mean value and the measured value at a location, and can be the result of errors (differences) due to analytical variability and/or to sampling variability. Typically, sampling variability (i.e., the chance that a soil sample drawn from a yard does not have a concentration equal to the true mean for that yard) is likely to be the largest source of measurement error.

The difficulties introduced by measurement error can be seen by inspecting the graphs shown in Figure 1 and Figure 2. Figure 1 displays a simulated data set of 400 data points generated using the following probability distribution functions:

C_s = Lognormal (1000,1000)
k_0 = Lognormal(500,500)
k_s = Normal(0.4,0.1)
$C_d = k_0 + k_s*C_s$

Despite the fact that each of the independent parameters is allowed to vary over a wide range, there is still a fairly clear relationship between Cs and Cd, and simple linear regression analysis yields good estimates for both k_0 and k_s:

k_0 = 514 (theoretical = 500)
k_s = 0.39 (theoretical = 0.40)

Figure 2 shows a simulated data set of 400 data points generated exactly as above, except the simulated values for C_s and C_d were each multiplied by a "measurement error" term defined as:

Measurement Error = Lognormal(1.0,0.6)

As seen, the data now show less of a clear pattern, and a simple linear regression analysis of the data tends to underestimate the value of k_s and overestimate the value of k_0:

k_0 = 620 (theoretical = 500)
k_s = 0.26 (theoretical = 0.40)

This pattern (underestimation of k_s and overestimation of k_0) becomes significant as the standard deviation of the measurement error becomes larger than about 0.3-0.4, and progressively worsens as the standard deviation of the measurement error increases. Thus, the first step in deciding how to utilize soil/dust data at a site is to evaluate the magnitude of the measurement error in the data, as discussed below.

Quantification of Measurement Error

The best way to evaluate measurement error is to compare the results of independent samples of soil taken from the same location (typically a yard). Note that splits (divided portions of the same sample) may be helpful in estimating analytical variability, but will not reveal sampling variability. Since sampling variability is likely to be the primary source of measurement error, it is best to collect at least two independent samples from all yards (or at least a number of yards). Analyze the measurement error as follows:

1. Calculate the mean value (M) for the contaminant in soil at each location. Then calculate the ratio (R) of the first measurement (X) over the mean for all sample pairs:

$$R_i = X_i/M_i$$

2. Calculate the mean and standard deviation of the ratios. It is expected that the mean will be close to 1.0. (If not, there is probably some systematic difference between the independent measurements collected). The standard deviation of the ratios is a good indicator of the size of the measurement error. Standard deviations less than about 0.3 are consistent with small error, while values of 0.5 or above indicate large measurement error. Values between 0.3 and 0.5 are intermediate in magnitude.

Approach When Measurement Error is Significant

Simulations run using a variety of assumptions about the distribution of C_s, k_0 and k_s indicate that the mean values for dust ($\bar{D}$) and soil ($\bar{S}$) are rather insensitive to the size of the measurement error, and so the ratio of mean dust over mean soil ($\bar{D}/\bar{S}$) is also rather stable and insensitive to measurement error. For example, the following results in Table 2 were obtained using the probability distribution functions given above:

Table 2

Measurement Error	$\bar{D}$	$\bar{S}$	$\bar{D}/\bar{S}$
Lognormal(1,0.0)	898	996	0.90
Lognormal(1,0.2)	900	1003	0.90
Lognormal(1,0.4)	899	973	0.92
Lognormal(1,0.6)	887	1041	0.85
Lognormal(1,0.8)	898	1064	0.84
Lognormal(1,1.0)	936	957	0.98

Note, however, the simple ratio of $\bar{D}/\bar{S}$ will overestimate k_s by a factor related to the size of k_0. Thus, if a reasonable value for k_0 can be estimated, it is simple to calculate an estimate for k_s that is unlikely to be seriously biased by measurement error using the following equation:

$$k_s \approx \frac{\bar{D} - k_0}{\bar{S}} \qquad (4)$$

where:

k_s = Mean mass fraction of soil in dust (unitless)

$\bar{D}$ = Mean concentration of all dust measurements (ppm)

k_0 = Mean concentration of contaminant in dust that is not attributable to yard soil (ppm)

$\bar{S}$ = Mean concentration of all soil samples (ppm)

If you have evidence (or reasonable basis to believe) that there are no significant interior sources, and that there are no significant areas sources (blowing waste piles, active industrial releases to air, etc.), then assume k_0 is equal to zero and estimate k_s as:

$$k_s \approx \frac{\bar{D}}{\bar{S}} \qquad (5)$$

If you suspect that there are either interior sources and/or area sources that contribute to measured interior dust measurements, then estimate k_0 by finding the mean of all dust samples from locations where the level in the outdoor soil is low (e.g., less than 20% of the mean). Simulations have shown that this approach usually gives a reasonable estimate for k_0 if there are enough data points, although the value is often biased slightly high. If you feel that this approach does not yield a reliable estimate of k_0 (for example, if your estimate of k_0 is nearly as large or larger than $\bar{D}$), consult with a professional statistician to determine how best to estimate k_0. If you do not believe that you can derive a reasonable estimate for k_0, then calculate k_s as in Equation 5. Be aware that this approach will overestimate k_s if k_0 is not close to zero.

Effect of Sample Size

Because the relationship between the concentration of contaminant in soil and that in indoor dust is expected to be highly variable, and because of the confounding effect of measurement error, it is very difficult to derive reliable estimates of either k_0 or k_s when there are fewer than several hundred measurements. The following table (Table 3) illustrates how the ability to estimate reliable values for these parameters deteriorates as sample size decreases or as measurement error increases. As before, the nominal estimates are $k_0 = 500$ and $k_s = 0.40$. Inspection of this table reveals that data sets as small as 50 may be useful if measurement error is small enough to permit use of linear regression techniques, although even this

Table 3

Sample Size	Method	Error = LN(1,0)		Error = LN(1,0.2)		Error = LN(1,0.6)	
		k_0	k_s	k_0	k_s	k_0	k_s
400	Regression	514	0.39	534	0.37	619	0.26
	Ratio	494	0.41	533	0.37	474	0.40
200	Regression	519	0.36	532	0.32	675	0.18
	Ratio	396	0.47	491	0.35	600	0.26
100	Regression	454	0.36	565	0.30	734	0.11
	Ratio	294	0.52	547	0.32	620	0.21
50	Regression	469	0.36	468	0.46	777	0.13
	Ratio	304	0.52	806	0.14	855	0.05

estimation technique tends to become rather unstable with this few data points. If measurement error is significant, then linear regression does not yield reliable results for any sample size, and the ratios method begins to yield poor results with sample sizes as small as 100-200, and yields clearly unreliable results with a data set of 50.

It is important to recognize that if the value of k_0 is small, then the simple ratio of $\bar{D}/\bar{S}$ will not be far from the true value of k_s. This is illustrated in Table 4 below. Thus, if k_0 is small (or if a decision is made to accept the conservatism introduced by assuming k_0 is small even if it is not), the number of samples needed is dictated only by the need to generate stable estimates of $\bar{D}$ and $\bar{S}$. Typically, this will be only about 20-30, depending of the variability between measurements.

Example Analysis (Leadville)

Figure 3 shows a set of 200 paired soil lead/dust lead data points collected at the California Gulch Superfund site in Leadville, Colorado. For each of these

Table 4

Error	k_0	$\bar{D}$	$\bar{S}$	$\bar{D}/\bar{S}$
Lognormal(1,0.6)	Lognormal(500,500)	888	956	0.93
Lognormal(1,0.6)	Lognormal(400,400)	795	971	0.82
Lognormal(1,0.6)	Lognormal(300,300)	691	974	0.71
Lognormal(1,0.6)	Lognormal(200,200)	605	1001	0.60
Lognormal(1,0.6)	Lognormal(100,100)	479	946	0.51
Lognormal(1,0.6)	0	422	1030	0.41

locations, there were at least two independent measurements of lead in yard soil. The mean of the ratios for perimeter soil compared to the yard average is 1.1, and the standard deviation for the ratios is 0.44. Based on the decision rule above, this degree of measurement error is likely to introduce a significant bias into estimates of k_0 and k_s derived using linear regression techniques, so the ratio method was employed. The results were as follows:

$\bar{D}$ = 839 ppm

$\bar{S}$ = 1333 ppm

$k_0 \approx 517$ (average of all dust measurements where soil lead is less than 200 ppm)

$k_s \approx 0.24$

The line in Figure 3 shows the values predicted by this model.

Other Sites

The importance of measurement error has only recently been appreciated, so most prior analyses have utilized simple linear regression techniques or structural modeling to estimate the value of k_s. Some example results, all based on lead measurements, are summarized below in Table 5.

Table 5

Site	k_s (ppm per ppm)	Reference
Sandy, UT	0.10	Unpublished analysis (WESTON)
Butte, MT	0.24	USEPA 1994
Kellogg, ID	0.09	Marcus and Bernholic 1990
Sharon Steel, UT	0.76	Marcus and Bernholic 1990
East Helena, MT	0.88	Marcus and Bernholic 1990

As seen, the value of k_s estimated in these analyses is significantly less than 1.0 for most of these sites, supporting the view that interior dust is not composed entirely of exterior soil. It is interesting to note that there is a substantial variation in the values between sites. This indicates that the amount of soil entering homes as dust varies between the sites, presumably due to differences in environmental conditions and/or differences in human activity and demographic factors. This inter-site variability emphasizes the need for site-specific data in evaluating esposure from dust.

RECOMMENDATION

Planning a Data-Collection Effort

As discussed above, the data needed to support a deviation from the default assumptions are a set of paired measurements of contaminant concentration in soil and dust from a series of different residences or buildings. Because of the wide variability typically observed in these data, it is highly desirable to collect samples from a large number of different locations (at least several hundred) in order to obtain a reliable estimate of the relation between soil and dust. Smaller data sets are likely to be useful only if the correlation is strong (i.e., variability is low) or if measurement error is small. Because of the potential problems introduced by measurement error, it is necessary to collect data that allow estimation of the magnitude of the measurement error. As discussed above, this is done by collecting at least two independent random samples of soil from exterior yards. Ideally, two or more samples should be collected from all yards, but multiple samples from a minimum of 30 randomly selected yards will usually be adequate.

Decision Tree

Figure 4 summarizes the basic decision process needed to evaluate contaminant concentration in interior dust. Each of the main branches of the tree are discussed below.

No Data

If you do not have site-specific data, use the following default assumptions when estimating the concentration of contaminants in indoor dust:

(a) Contaminants other than lead,
$C_d = 1.0 * C_s$

(b) If lead is the contaminant,
$C_d = 0.7 * C_s$

Use the uncertainties section to discuss the likelihood that the default assumptions may be overconservative, and you may a provide quantitative discussion of how much the default assumption might overestimate risk if the soil-dust ratio were lower than the defaults.

Site-Specific Data

If you have site-specific data (paired measurements of contaminant in soil and dust), the approach you follow will depend on a) how many data points you have, and b) the magnitude of the measurement error in your data. The basic rules are as follows:

- Use linear regression only if you know that measurement error is small. This approach will probably yield reliable results with as few as about 50 data points, but 100-200 (or even more) will provide for greater reliability.

- If you find that measurement error is large, or if you don't know the size of the measurement error, assume that you must use the ratio method. If you have a large number of samples, estimate k_0 by stratifying the data and calculating the mean value of C_d for all samples where C_s is small, then calculate k_s using Equation 4:

$$k_s \approx \frac{D - k_0}{S} \qquad (4)$$

If you cannot reliably estimate k_0, then estimate k_s using Equation 5:

$$k_s \approx \frac{D}{S} \qquad (5)$$

As noted above, this approach will tend to overestimate the true mean value of k_s by a factor related to the size of k_0. If the resulting estimate is larger than the defaults, use the defaults instead.

DATA APPLICATION TO RISK CALCULATIONS

There are two alternative approaches for using site specific data to quantify dust exposure:

1) Utilize location-specific data for all locations where dust has been measured, and use the values of k_s and k_0 (derived as above) to estimate C_d from the value of C_s at all locations

where dust has not been measured.

2) Utilize the values of k_s and k_0 to calculate a value for C_d at all locations, including those where C_d has been measured.

Each approach has some advantages and some limitations. Usually, measured values are more reliable than modeled values, which favors selection of option 1. However, based on the assumption that the contribution of yard soil to house dust is not a constant but depends on human activity patterns which determine how much soil enters the house, then a measured value at a location is applicable only to the current occupants of the building at that location, and might overestimate or underestimate the soil contribution for some other future occupants of the building. In addition, measured values do not distinguish interior sources (e.g., paint) from exterior (soil-related) sources. Thus, use of measured values may yield an accurate estimate of total risk, but may tend to over-estimate soil-related risk.

In general, Region VIII recommends using approach 1 (calculation of C_d only at locations where direct dust measurements are not available). Consult with your regional toxicologist to determine if there is any reason to consider using approach 2.

Data Application to PRG Calculations

The basic approach for calculating a PRG for a contaminant in soil is to solve the risk equation for the concentration which corresponds to a specified target risk level. Typically, the exposure from soil is taken to include the intake of soil present in interior dust. Thus, the equation that relates concentration level in dust to that in soil is needed in PRG calculations as well as risk calculations.

Assuming that the value of k_0 (the concentration of contaminant in dust that is attributable to indoor and/or area sources) is greater than zero, it is a matter of risk management judgement or policy whether this contribution to total dust concentration should be included in the PRG calculation. That is, in the PRG equation, the equation used to estimate C_d from C_s could be either:

$$C_d = k_0 + k_s*C_s$$

or:

$$C_d = k_s*C_s$$

If the first equation is chosen, then the yard clean-up must be more stringent that it would be otherwise to account for the exposure from non-yard sources (those described by k_0). If the second equation is used, then the health-based goals for total exposure (including those exposures from non-yard sources described by the k_0 term) may not be met.

In general, Region VIII recommends that the second equation should be used. That is, the contribution of non-yard sources should <u>not</u> be included in calculation of yard clean-up goals. This recommendation should be discussed with the remedial project manager, the regional toxicologist, and possibly regional counsel before PRG calculations are performed and action levels are set.

REFERENCES

1. Bornschein, R. L., P. Succop, R. N. Dietrich, C. S. Clark, S. Que Hee, P. B. Hammomd. 1985. The Influence of Social and Environmental Factors on Dust Lead, Hand Lead, and Blood lead Levels in Young Children. Environ. Res. 38:108-118.

2. EPA. 1989. Risk Assessment Guidance for Superfund. Volume I. Human Health Evaluation Manual (Part A). Interim final. U.S. Environmental Protection Agency, Office of Emergency and Remedial Response. Washington, DC. Publication Number EPA/540/1-89/002.

3. EPA. 1991. Risk Assessment Guidance for Superfund. Volume I: Human Health Evaluation Manual. <u>Supplemental Guidance</u> "Standard Default Exposure Factors". Interim Final. U.S. Environmental Protection Agency, Office of Emergency and Remedial Response. Washington, D.C. Publication Number 9285.6-03.

4. EPA. 1993. Superfund's Standard Default Exposure Factors for the Central tendency and Reasonable Maximum Exposure. Draft. U.S. Environmental Protection Agency, Office of Emergency and Remedial Response. Washington, D.C.

5. EPA. 1994. Guidance Manual for the Integrated Exposure Uptake Biokinetic Model for Lead in Children. U.S. Environmental Protection Agency, Office of Emergency and Remedial Response. Washington, D.C. Publication Number 9285.7-15-1.

6. Lanphear B.P., M. Weitzman, M. Tanner, T. Clarkson, N. L. Winter, M. Emond, B. Yaakir, A. Watts, S. Eberly, L. Apetz, T. Rood. 1994. The Relation of Lead-Contaminated House Dust and Blood-Lead Levels Among Urban Children. Department of Pediatrics, Biostatistics, and Environmental Medicine, University of Rochester School of Medicine and Dentistry. Draft Final Report submitted to the U.S. Department of Housing and Urban Development, National Center for Lead-Safe Housing.

7. Marcus A.H., A. Bernholic. 1990. Inter-Site Comparisons of Environmental Lead Uptake. Battelle Statistics and Data Analysis Systems, Research Triangle Park, NC.

8. Weston. 1995. Draft Final Baseline Human Health Risk Assessment for the California Gulch Superfund Site. Prepared by Roy F. Weston, Inc. for U.S. EPA Region VIII. Lakewood, Colorado.

Figure 1 Simulated Soil/Dust Data

No Measurement Error

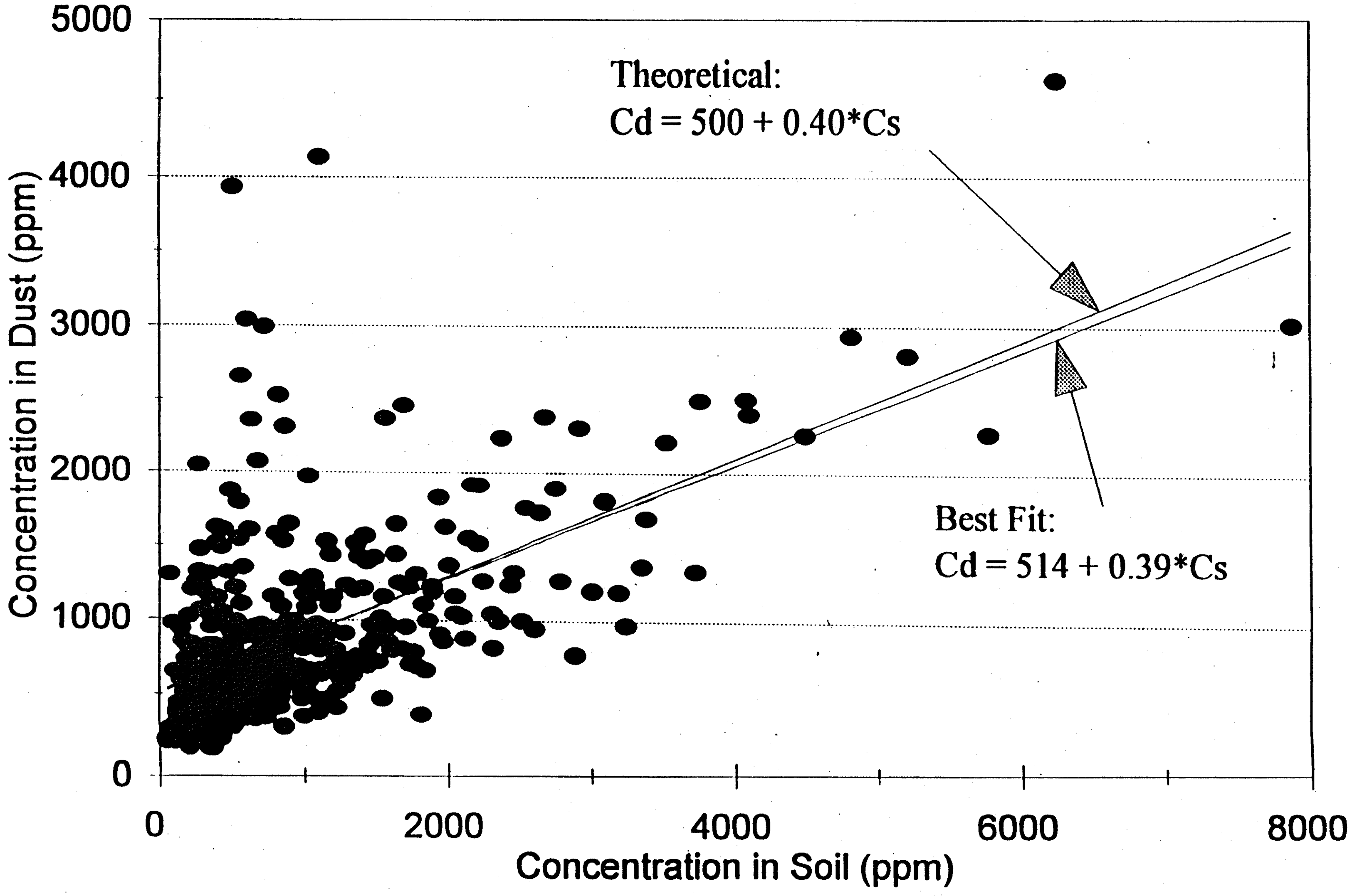


Figure 2 Simulated Soil/Dust Data
With Measurement Error

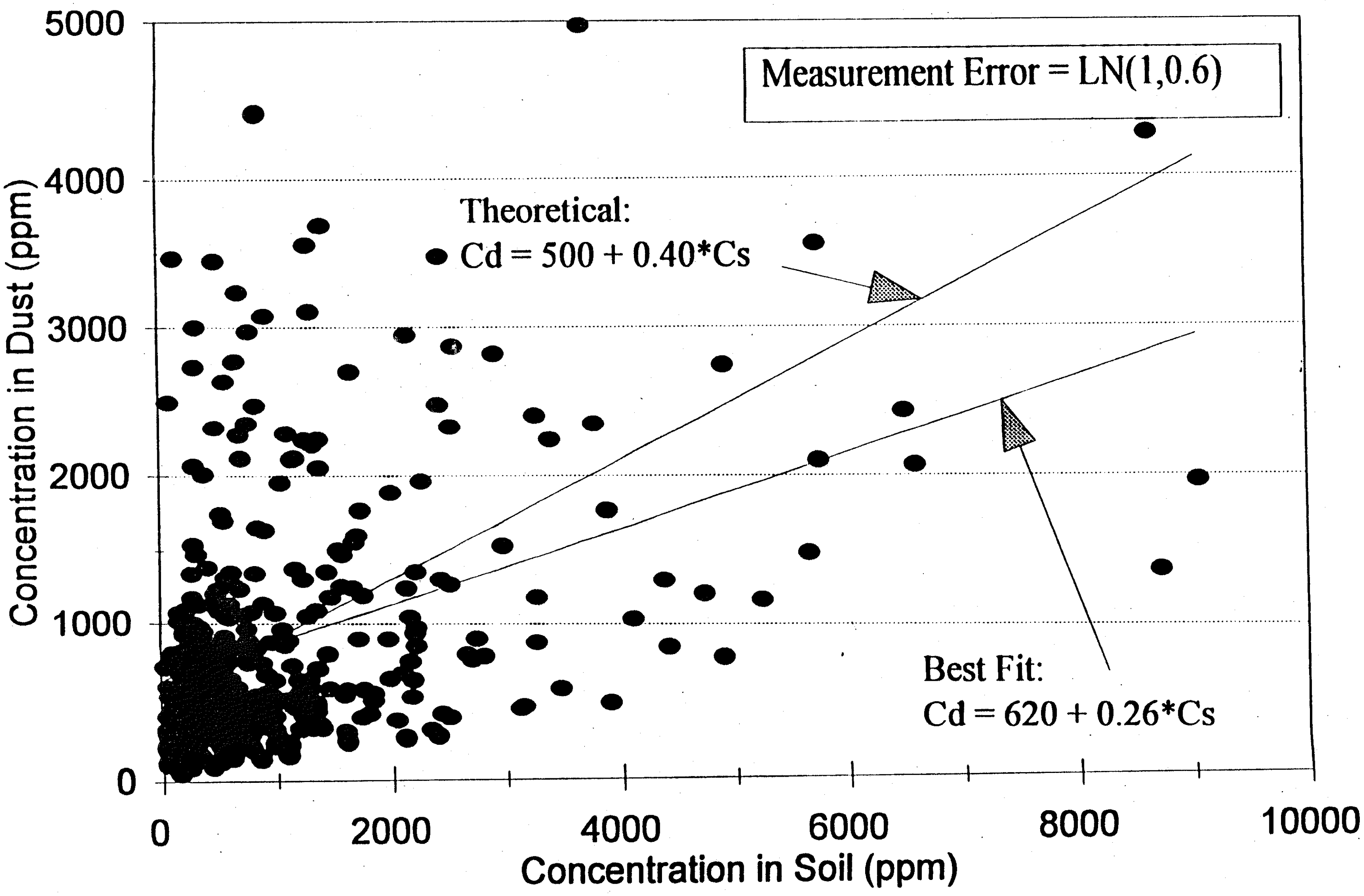
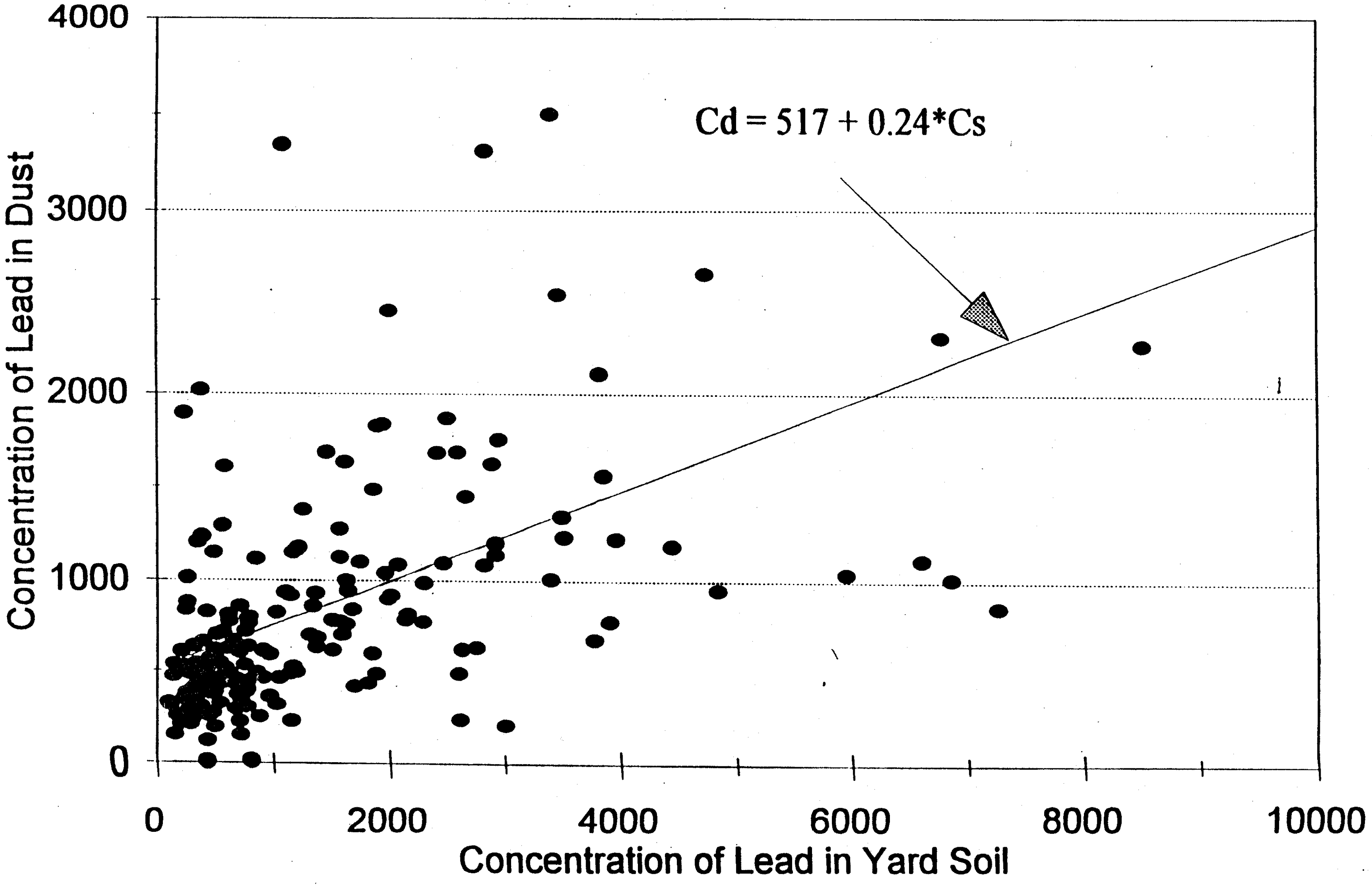


Figure 3 Soil/Dust Data from Leadville



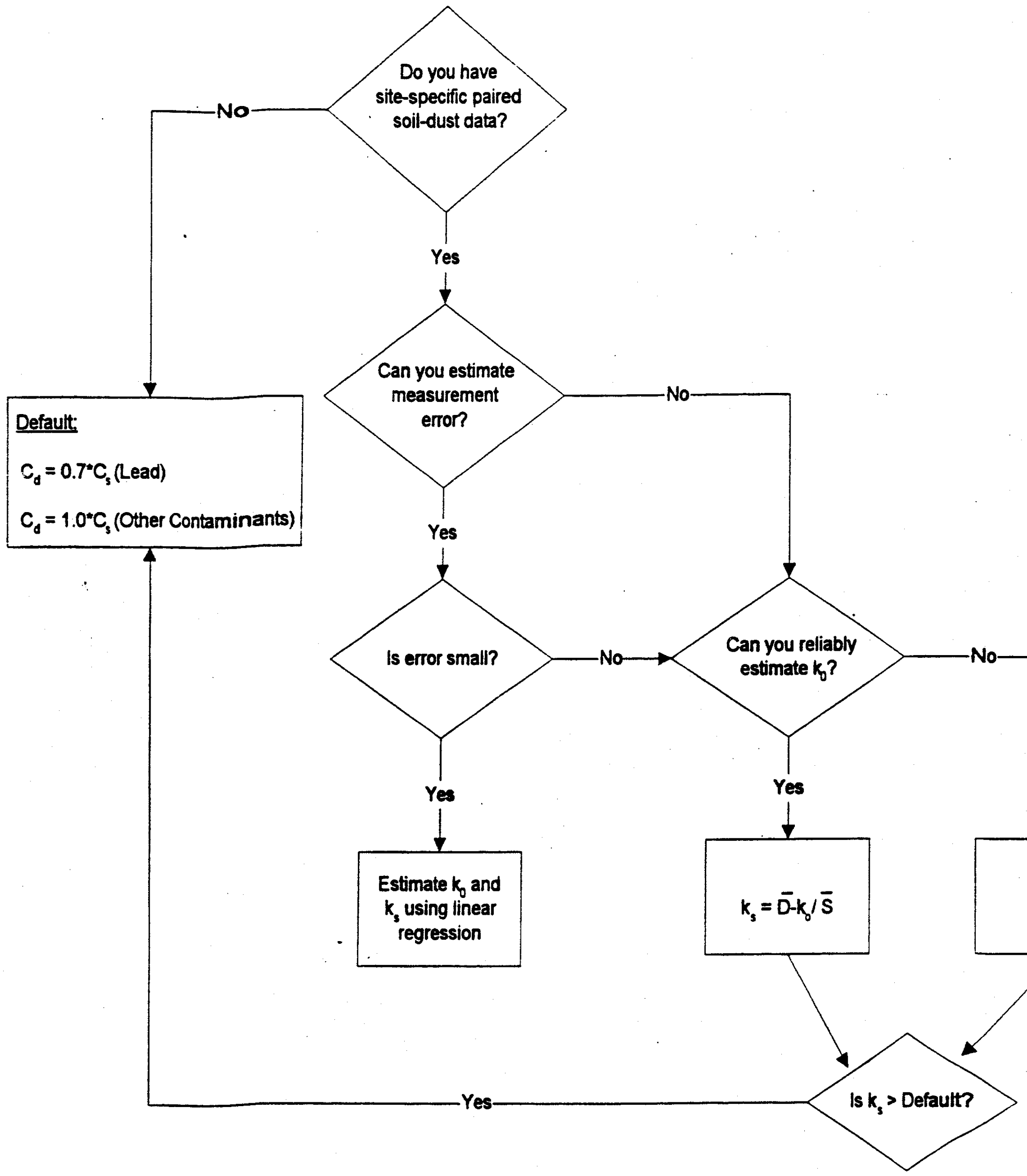


Figure 4 Decision Tree