United States
Environmental Protection
Agency

Office of Pollution
Prevention and Toxics
(7401)

EPA 747-R-93-011
September 1993

Pollution Prevention and Toxics

EPA

Data Analysis of Lead In Soil and Dust

Data Analysis of Lead In Soil And Dust

This work was conducted under contract number
68-D9-0174, 68-D2-0139, and 68-D3-0011

Prepared for
Samuel Brown, Work Assignment Manager
Technical Programs Branch
Chemical Management Division
Office of Pollution Prevention and Toxics
U. S. Environmental Protection Agency
Washington, D.C. 20460

September 1993

The material in this document has been subject to Agency technical and policy review and approved for publication as an EPA report. The views expressed by individual authors, however, are their own and do not necessarily reflect those of the U.S. Environmental Protection Agency. Mention of trade names, products, or services does not convey, and should not be interpreted as conveying, official EPA approval, endorsement, or recommendation.

Contents

Tables

Figures

Acknowledgments

The Office of Pollution Prevention and Toxics of EPA would like to express their appreciation for the many efforts and the contributions of Westat in the data analysis, interpretation, writing, and preparation of this report. We would also like to thank, Phil Robinson, John Schwemberger, Brad Schultz, and Ben Lim for their guidance and support throughout this research.

Executive Summary

Purpose

The 1987 amendments to the Lead-Based Paint Poisoning Prevention Act required the Department of Housing and Urban Development (HUD) to prepare and transmit to Congress "an estimate of the amount, characteristics and regional distribution of housing in the United States that contains lead-based paint hazards." In response to this mandate, HUD sponsored a national survey of lead-based paint in housing. HUD's *Comprehensive and Workable Plan for the Abatement of Lead-Based Paint in Privately-Owned Housing: A Report to Congress* and EPA's *Report On The National Survey Of Lead-Based Paint In Housing* documents the survey and presents considerable data on the extent and characteristics of the lead paint hazard in homes. The purpose of this study is to supplement the prior reports through additional data analyses focusing on the contribution of lead in exterior soil to the lead hazard in homes.

The specific objectives of this study are two-fold: First, to conduct analyses of the statistical associations among the soil, dust, and paint data from the national survey. Many researchers believe that lead contamination in soil originates mainly from paint lead and automobile emissions. Similarly, interior dust lead is believed to come principally from paint lead and soil lead. The objective is to explore these hypotheses through a statistical analyses of the national survey data. Second, to present an analytical description of the soil and dust data, including sampling and measurement errors in the data, in order to evaluate the suitability of the soil and dust data for future data analyses.

The statistical analyses reported here focus on techniques that identify the variables that "best predict" soil lead and dust lead levels. It is to be noted that a strong statistical association among these variables does not, by itself, prove that one variable represents a source of the lead measured by another, e.g., that exterior soil lead is a source of interior dust lead. Both soil and dust lead contamination levels may be caused by a third source of lead such as paint lead, or automobile emissions.

National Survey Methodology

The study population for the national survey consisted of nearly all housing units in the United States built before 1980. (Newer homes were excluded because the Consumer Product Safety Commission banned the use of lead-based paint in residences in 1978.) The national survey was conducted in 381 housing units, 284 privately-owned and 97 public housing units, selected to represent the entire pre-1980 United States housing stock. This report focuses on the soil and dust data from privately owned housing.

In each sampled housing unit, one room with plumbing ("wet") and one room without plumbing ("dry") were randomly selected for inspection. Painted surfaces were inventoried and measured, and their conditions assessed. An exterior wall was similarly selected and inspected. Paint lead measurements were obtained *in situ* on interior and exterior surfaces using portable X-ray fluorescence (XRF) spectrum analyzers. The analyzers

measured area concentrations, described as milligrams of lead per square centimeter of painted surface (mg/cm^2).

Soil samples were collected at three exterior locations on the property of each dwelling: (1) a *drip-line* sample near an exterior wall of the dwelling, potentially contaminated with deteriorated lead-based paint; (2) an *entrance* sample collected near the most commonly used entrance, to measure the potential associations with track-in lead; and (3) a *remote* sample, intended to measure background lead from sources other than lead-based paint.

Dust samples were collected with a vacuum sampler at seven interior locations within each home: from floors, window wells (the place at the bottom of the window where the sash rests when it is closed), and window sills (the part inside the room) in the sampled wet and dry rooms; and from the floor just inside the main entrance to the housing unit. Lead loading levels (area concentrations) were determined--micrograms of lead per square foot of surface ($\mu g/ft^2$), along with mass concentrations of dust lead in ppm.

Soil and dust samples were sent to laboratories and analyzed by inductively coupled plasma-atomic emission spectrometry (ICP-AES) and by graphite furnace atomic absorption (GFAA) spectroscopy, respectively.

Sources of Soil Lead

Equations were developed to statistically relate the soil lead levels to a number of potential sources of soil lead and related factors, including exterior and interior paint lead loadings, percentage of damaged paint, and surface areas covered with paint; dwelling unit age and other descriptors of the housing unit; number of rooms, local traffic volumes, county of residence, and 1920-1990 decennial Census populations. Combined, these potential sources of soil lead account for over 50 percent of the statistical variation in the lead in soil data.

The most significant predictors of soil lead concentrations at all three soil sample locations are dwelling unit age and county of residence. Soil lead concentrations increase with dwelling unit age. Dwelling unit age measures the length of time since the construction of the building and, in most cases, the last major disturbance of the soil. Thus, dwelling unit age measures the length of time lead deposits -- from whatever source -- have been accumulating on the soil. The county of residence effect may be due to many factors including regional variations in population density, population growth, background soil lead levels, traffic, and home building and painting practices. Local traffic volumes are significantly related to soil lead at the remote and drip line locations.

The soil lead concentrations at all three sampling locations were closely related to the overall dwelling unit paint lead loadings (represented by the geometric mean of the paint lead loadings on the wet room, dry room, and exterior surfaces). The results suggest that paint lead from dwelling surfaces contribute more to the entrance and drip line soil lead samples than to the remote sample. This finding was expected because entrance and drip-line samples are closer to painted structures than are remote samples.

While both interior and exterior paint lead relate significantly to soil lead, exterior paint lead is more strongly associated with soil lead than is interior paint.

Sources of Dust Lead

The statistical relationships were studied between interior dust lead levels -- for all seven dust sample locations -- and a number of possible sources of dust lead and related factors, including housing unit paint lead loadings (the area weighted average across all painted surfaces), percentage of damaged paint, and surface area covered with paint; dwelling unit age and other descriptors of the housing unit; and all three soil samples. Generally, the dust lead has more variation than the soil lead data. This makes it more difficult to identify and assess significant sources of dust lead. The dust lead equations account for only 16 to 27 percent of the statistical variation in the dust lead data. The findings regarding sources of dust lead are therefore more tentative and less conclusive than those regarding the sources of soil lead. Nevertheless, some significant factors relating to dust lead have been identified.

Floor dust lead at the main entrance is statistically associated primarily with exterior soil lead and, to some extent, exterior paint that is both leaded and damaged. It appears that the soil lead contribution comes mainly from the entrance soil samples. However, both close-in soil sample locations (entrance and drip line) contribute to dust lead concentrations just inside the main entrance.

Floor dust lead in the wet and dry rooms appears to come more from soil lead at the two close-in locations than directly from paint lead. That is, while there is clear evidence of a statistical association between soil lead and floor dust lead, the evidence is less clear of a direct association between floor dust lead and paint lead. There is one exception to this: floor dust lead in the wet room is significantly associated with wet room paint lead. However, as described in the previous section, soil lead is related to overall paint lead. This suggests that, over time, lead migrates from paint, to soil, to floor dust.

Soil lead concentrations at the close-in locations are significant predictors of dry room window sill dust lead concentrations. Interior, but not exterior, paint lead is also associated with dry room window sill dust. Wet room window sill dust lead is significantly related to interior paint lead, especially in the wet room.

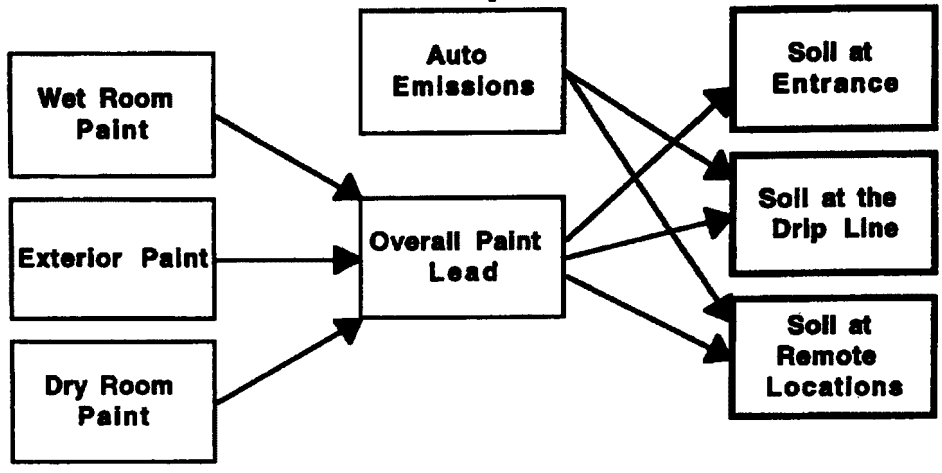
There were fewer window well dust samples to analyze and these are the most variable of the dust samples; consequently, the statistical analyses do not permit any assessment of the sources of wet or dry room well dust lead levels. However, the fact that dust lead concentrations in window wells are significantly higher than soil lead concentrations suggests that other sources, such as paint, contribute much more lead to window well dust than does soil.

Note that the dust lead concentrations increase significantly with age of the dwelling unit, as do the soil lead concentrations. However, in the statistical analysis, age was not significant in predicting dust lead concentrations, indicating that other factors that were also related to age, such as soil lead concentrations and paint lead loadings, were adequate to predict the dust lead concentrations without an additional age term.

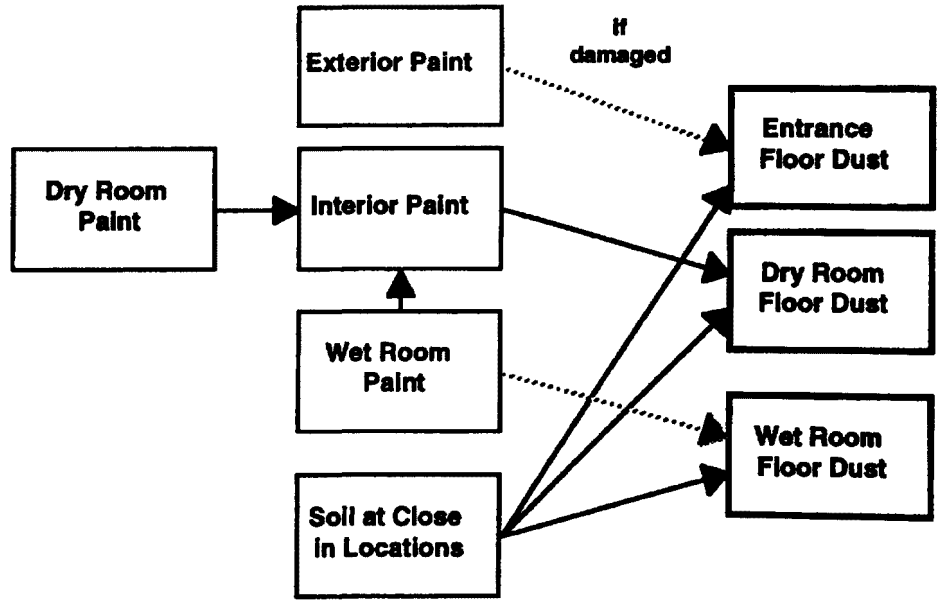
Pathways

The relational analyses described above suggest certain conclusions concerning the pathways by which lead migrates from paint, automobile emissions, and other sources to exterior soil and interior dust. These conclusions are summarized in the following diagrams of identified pathways of lead from these sources to soil and floor dust. The diagram shows only pathways identified as significant in the analysis of the national survey data. Additional pathways, not identifiable from the national survey data, may exist. The solid arrows indicate clearly identified paths; the dotted lines indicate paths for which the evidence in the national survey data is less clear.

Identified Pathways to Soil Lead



Identified Pathways to Floor Dust Lead



Soil Lead Data Analysis

The soil lead data for each sample location can be statistically described by a log normal distribution (i.e., the logarithm of the measured lead concentrations have a normal or "bell-shaped" distribution). No individual observations can be clearly identified as outliers. Therefore, all of the soil measurements are included in the analyses in this report.

The arithmetic means for the entrance, drip-line, and remote samples are 295, 415, and 170 ppm, respectively; the geometric means are 83, 72, and 47 ppm, respectively. While the measurement error is relatively large (about 95 percent of soil lead measurements will be within a factor of 2.7 of the true concentration), it is small compared to the differences in soil lead concentrations between locations. In spite of the measurement error, significant correlations and differences can be found. In particular, the lead concentrations at the entrance and drip-line locations are, on average, significantly higher than at the remote sample location ($p < 0.001$). However, the entrance and drip-line samples are not significantly different from each other. The measurements at the three locations are also all highly correlated with each other; that is, housing units that have higher (lower) lead concentrations at one of the locations also tend to have higher (lower) concentrations at the other two. The soil lead measurements also vary significantly with the age of the dwelling unit. Homes built before 1940 have, on average, the largest soil lead concentrations; after 1940, the average concentrations decline with each successive decade.

Dust Lead Data Analysis

The dust lead data can also be statistically described by a log normal distribution for each sample location. About one percent of the individual observations can be clearly identified as outliers -- observations that are extremely unusual when compared to other observations made under comparable conditions. These observations have been removed from the analyses reported here because they could have obscured and distorted the relational analyses. In general, the dust lead data is noticeably more variable than the soil lead data, with the window locations more variable than the floor locations.

On average, the wet and dry room floor dust levels are similar to each other, while the entrance way dust lead level is significantly greater. The difference may represent what is tracked into the house from outside. All of the window dust lead levels are significantly greater than any of the floor lead levels. Further, the window well lead levels are usually greater than the lead levels on the same window sills. There are no significant differences between corresponding locations in the wet and dry rooms. The measurements at the seven dust sampling locations are also all highly correlated with each other.

There is no one widely-accepted dust sample collection protocol. Consequently, researchers in different studies may use different methods; which means that data reported by two different researchers may not be comparable. In particular, some researchers use wet wipes to collect dust samples. There is evidence that wet wipe samples tend to yield higher dust lead loadings than the vacuum used in the national

survey. Consequently, caution must be taken in comparing these results with other studies.

Limitations Of The Analysis Results

Although appropriate for the objectives of the original survey, the data provided limited ability to identify possible sources of soil and dust lead. As a result of (1) the limited number of XRF measurements, soil measurements, and dust measurements, (2) the lack of information on the behavior of household occupants, and (3) inherent variability in the sampling and measurement process, the possibilities for identifying the sources of lead in soil and dust are limited and the conclusions are subject to interpretation. That the models cannot accurately predict lead concentrations in homes suggests that the lead levels in dust and soil are determined, in part, by factors that were not recorded in the survey or are difficult to quantify, such as small-scale local factors and the behavior of the occupants. Nevertheless, the statistical procedures identified some significant relationships in the data, support conclusions reached by other researchers, and provide valuable descriptive statistics for describing homes nationally, statistics that are not available from other sources.

Summary

The overall summary of the report's findings are presented below in the form of general questions that might be asked of the reports conclusions, and brief answers to these questions.

1) What are the main contributors of lead to soil?

   The analysis suggests that both paint lead and lead from traffic-related sources have contributed to soil lead, with traffic-related sources being more important at locations away from buildings and paint sources more important next to buildings, at the drip line and entrance. In addition, the soil lead concentrations increase significantly with increasing dwelling unit age due to factors that cannot be identified by this analysis.

2) Based on these analyses, are automobile emissions still a key contributor to soil lead? If so, how much?

   Although the results suggest that automobile emissions have contributed lead to soil, particularly in soil around older homes, the contribution of additional lead from automobile emissions today cannot be determined from the survey data.

3) How accurate are the conclusions based on the analysis?

   The statistical conclusions provide support to generally accepted hypotheses about the sources of lead in soil and dust. They provide some indication of the relative magnitude of the contribution from different sources. The results are not accurate enough to identify lead sources and pathways with confidence, to identify the less

important lead sources, or to predict dust and soil lead concentrations in individual homes.

4) Based on these analyses, what is the main contributor to dust lead inside the home, paint or soil?

The results suggest that, in general, soil is more important than paint as a source of lead in floor dust. The relative importance of paint as a source of lead in dust increases if the paint is damaged. However, the importance of paint damage cannot be reliably determined from the data. No conclusions can be reached about window sill and well dust; however, the high lead concentrations in window well dust suggest that paint, rather than soil or floor dust sources, contributes most of the lead to window well dust.

5) Are there any major findings of this report that could be used to combat our lead poisoning problem?

The analysis suggests that the dust lead concentration is statistically independent of the dust loading. Lead loading, thought to be most closely related to a child's risk of lead poisoning, can be expressed as the product of the lead concentration and the dust loading. Therefore reducing either the dust loading, perhaps by frequent vacuuming, or the lead concentration, perhaps by removing lead sources, can reduce lead loading and lead poisoning risk. The study suggests that soil dust lead is the major contributor of lead to floor dust. Therefore, effective measures to reduce the movement of soil into the home can also help control floor dust lead.

1. Introduction

1.1 Background And Objectives

The 1987 amendments to the Lead-Based Paint Poisoning Prevention Act required the Secretary of Housing and Urban Development (HUD) to prepare and transmit to Congress "a comprehensive and workable plan" for the abatement of lead-based paint in housing and "an estimate of the amount, characteristics and regional distribution of housing in the United States that contains lead-based paint hazards at differing levels of contamination." In response to this mandate, HUD sponsored a National Survey of Lead-Based Paint in Housing.

The National Survey of Lead-Based Paint in Housing produced a detailed, statistically valid, national database and study on the nature and extent of the potential hazards from lead-based paint. Soil and dust data were collected in order to support the analyses of the paint data. These data - paint, soil, dust - have been, and will continue to be analyzed to support the development of federal policy and programs with respect to the lead hazard in homes. These supportive analyses generally take the form of describing the problem, estimating its magnitude, identifying circumstances for priority government action, and estimating the costs and benefits of different policies. The specific objectives and analytic requirements of many of these analyses were not foreseen when the survey was designed and implemented. The suitability of the soil data for future, unanticipated analyses therefore needs to be determined. This includes an analyses of what the data actually represent, and of sampling and measurement errors in the data.

One issue currently before EPA involves the relationships between soil lead and paint lead. Exterior paint lead is believed to be a significant contributor to soil lead. In turn, soil lead is believed to be a significant contributor to the lead hazard in homes since children often come in contact with lead through soil and dust. The National Survey of Lead-Based Paint in Housing did not collect data on any direct measure of lead exposure, such as occupants' blood lead. However, it did collect data on dust lead, which is believed to be a major proximate source of blood lead, especially in children. An analysis of the relationships among soil, paint, and dust would estimate the potential hazard from soil and paint lead through dust lead.

The present study was designed and conducted to address the issues outlined above. The specific objectives of this study are as follows:

- To investigate the major sources of error in the National Survey soil data. These error sources include sampling error and laboratory analysis error. One goal is to estimate the potential impact of these error sources on classification biases in the estimates of the incidence of soil lead in housing.

- To statistically analyze the soil lead data, including measuring correlations between sampling locations within dwelling units.

- To statistically analyze the relationships among paint lead, soil lead, and dust lead, with a view toward estimating the contribution of soil lead to the potential lead hazard in privately owned homes, as measured by interior dust lead levels.

Soil lead analyses are reported here for privately-owned, occupied housing built before 1980. Post-1980 and vacant housing were outside the scope of the National Survey. While public housing was included in the National Survey of Lead-Based Paint in Housing, the public housing soil and dust data are somewhat suspect, as discussed in the *Report On The National Survey Of Lead-Based Paint In Housing.* Specifically, vacant public housing units were over-represented in the sample, and dust lead levels and incidence of damaged paint were lower in public housing than in private housing. In addition, many public housing units had all the property paved; there was no soil on the premises. Public housing data has therefore been excluded from this study. There were no similar problems associated with the data from private dwelling units.

Prior analyses of soil lead data from the National Survey of Lead-Based Paint in Housing appear in three reports: (1) the *Comprehensive and Workable Plan for the Abatement of Lead-Based Paint in Privately Owned Housing: Report to Congress*, prepared by HUD; (2) the *Report On The National Survey Of Lead-Based Paint In Housing*, prepared for EPA and composed of three volumes; and (3) the *Analysis of Soil and Dust Samples for Lead (Pb)*, also prepared for EPA.

These prior analyses indicate that an estimated 18 percent of privately-owned housing units have soil lead levels in excess of 500 ppm (although there is no federal standard for residential soil lead contamination, many experts agree that 500 ppm is a feasible threshold to designate "high" soil lead contamination in residential environments). EPA's interim guidance on soil lead cleanup levels at Superfund sites sets the cleanup levels at 500 to 1000 ppm[1]. The prior analysis also indicates that a similar 18 percent of homes have dust lead loading levels in excess of the HUD federal action levels as reported in the HUD Interim Guidelines for hazard identification and abatement of lead-based paint in housing[2]--200 μg/sq ft on floors, 500 μg/sq ft on window sills, and 800 μg/sq ft on window wells. (One important caveat, however, should be noted when comparing dust loadings measured in the National Survey to the HUD guidelines. In the national survey, dust was collected with a vacuum sampler, while the HUD guidelines are based on dust wiped from surfaces with wet wipes. Relationships between different dust collection techniques are not fully studied). The analyses also found a statistical association between the presence of damaged lead-based paint ("lead-based paint" is defined by HUD

[1]U.S. Environmental Protection Agency (September 7, 1989), Interim Guidance on Establishing Soil Lead Cleanup Levels at Superfund Sites (OSWER Directive # 9355.4-02).

[2]U.S. Department of Housing and Urban Development (1990), "Lead Based Paint: Interim Guidelines for Hazard Identification and Abatement in Public Housing", *Federal Register* 55 (April 18): 14557-14789.

as containing a paint lead concentration of 1.0 mg/sq cm or greater) and elevated soil and dust lead levels. This is a significant finding in light of potential exposures to children.

1.2 National Survey Methodology

This section presents a brief description of the objectives and methodology of the national survey of lead-based paint in housing. The objective of the national survey was to obtain data for estimating: (1) the number of housing units with lead-based paint; (2) the surface area of lead-based paint in housing and the associated estimate of national abatement costs; (3) the condition of the paint; (4) the incidence of lead in dust in housing units and of lead in soil near residential structures; and (5) the characteristics of housing with varying levels of potential hazard to examine possible priorities for abatement.

The study population consisted of nearly all housing in the United States constructed before 1980. Newer houses were presumed to be lead-free because, in 1978, the Consumer Product Safety Commission banned the sale of residential paint containing more than 0.06% lead by weight to consumers. The survey was conducted between December 1989 and March 1990 in 30 counties across the 48 contiguous states, selected to represent the entire United States housing stock, both public and privately-owned. The total sample size is 381 dwelling units, 284 privately owned and 97 publicly owned. The sample was small, but it provided estimates on the incidence of lead-based paint in housing that were sufficiently precise to develop the Congressionally-mandated *Comprehensive and Workable Plans* for Private and Public Housing.

Within each housing unit, two rooms were randomly selected for inspection, one with plumbing ("wet" room) and one without plumbing ("dry" room). In each room the field technicians inventoried painted surfaces, measured their dimensions, and assessed the condition of the paint. They also measured the lead concentration on randomly selected painted surfaces with portable "spectrum reading" X-Ray fluorescence (XRF) analyzers.[1]

Since not all rooms in a dwelling unit were inspected, it was possible to not detect lead-based paint when it was really present elsewhere in the dwelling unit. To reduce the chances of misclassifying a dwelling unit with lead-based paint as lead-free, additional lead readings, termed purposive readings, were taken on surfaces that, in the opinion of the field technicians, were most likely to have lead-based paint. Purposive measurements were not limited to the previously inspected wet and dry rooms. In some dwelling units, these additional purposive samples did, indeed, find lead-based paint where none had been found in the randomly selected rooms.

Exterior painted surfaces of each dwelling unit were also inventoried and XRF measurements were made on one randomly selected side of the house to detect the presence of lead in paint. Similarly, common areas within multifamily units (e.g., hallways, stairs, laundry rooms) were sampled and inspected.

---

[1]Details of the calibration and performance of the XRF analyzers can be found in *Report On The National Survey Of Lead-Based Paint In Housing.*

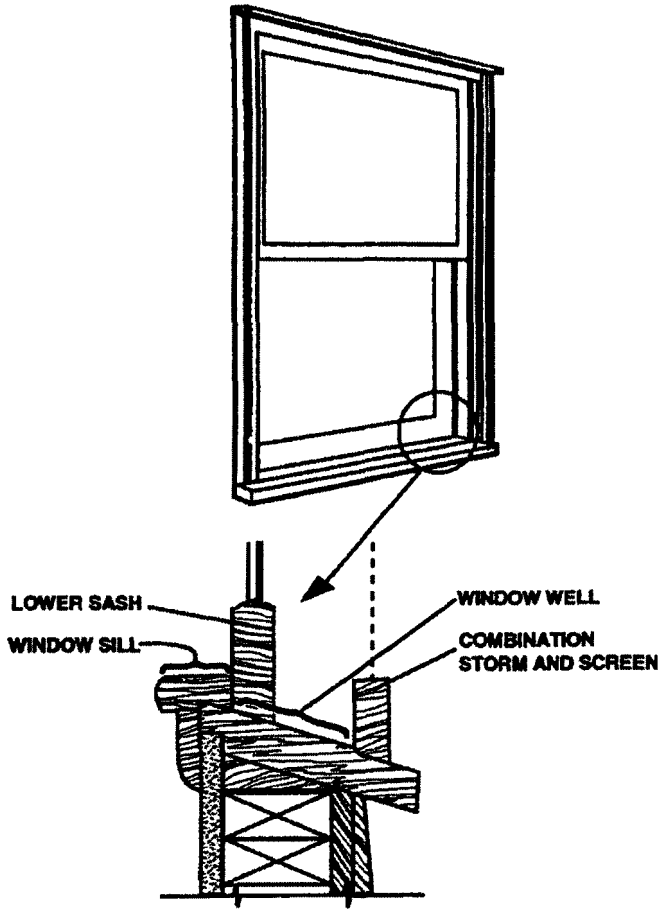
Exterior soil samples and interior dust samples were also collected. Generally, three soil samples were taken from each dwelling unit: (1) outside the main entrance to the building, (2) along the drip line of the sampled exterior painted surface, and (3) at a remote location away from the building but still on the property. Both the drip-line and the remote samples were normally collected on the same, randomly selected side of the house as the XRF paint lead measurement was made. Dust samples were collected on floors, window wells, and window sills in the wet and dry rooms and from the floor just inside the main entrance to the dwelling unit. Window wells are defined by HUD as the bottom of a window between the screen and the glass. Window sills are the lower part of the window inside the room.[1] Figure 1-1 illustrates the definitions of the terms "window well" and "window sill" as used in the National Survey of Lead-Based Paint in Housing. In cases where rooms had more than one window, field technicians chose one window and collected a sample. If insufficient dust was collected from sampling only one window, as determined by comparing filters to photographs of properly loaded filters, then another window in the same room was sampled into the same filter (only one sample for window wells and sills were obtained from each "wet" and "dry" room). Dust samples were also collected from common areas inside private multifamily housing units. However, since the sample size is small, they are not included in this report. Both dust and soil samples were sent to laboratories for lead analysis.

1.3 Organization Of This Report

This report has five chapters, after this introductory chapter. Chapters 2 and 3 present detailed descriptions of the soil lead data, including how it was collected, what it represents, and basic descriptive statistics. Included in Chapter 3 is an exploratory data analysis to identify and document outliers, censored data, and other anomalous aspects of the data that may affect future analyses and their interpretations. Chapter 4 presents a summary of the dust lead data. The summary focuses on evaluating the suitability of the dust lead data for the analyses reported in this report. Chapter 5 presents an analysis of the classification bias that results when one classifies a housing unit as having soil lead above a particular value, based on one to three soil lead observations. Finally, Chapter 6 presents the methodology and results of an analysis of the statistical association between dust lead, soil lead and paint lead, based on the data from the National Survey of Lead-Based Paint in Housing.

[1]Definitions in the HUD guidelines would be more consistent with common carpentry terminology if the term "window sill" was changed to "window stool" and the term "window well" was changed to "window sill".

Figure 1-1 Illustration of carpentry terms used in the report



2. Sampling And Lab Procedures For Soil Data

This chapter presents a background description and evaluation of the soil lead data collected in the National Survey of Lead-Based Paint in Housing. The chapter begins with a qualitative description of the field protocols used to collect the soil samples and the laboratory protocols used to analyze them for lead content.

2.1 Field Protocols

This section describes the protocols developed for collecting soil samples in the National Survey. The following information was drawn from the training materials provided to technicians and interviewers.

Technicians took three or four composite soil samples around each dwelling unit. The samples were taken from locations (1) as close as possible to the most frequently used **entry way** to the dwelling unit, or to the building housing the dwelling unit; (2) along the **drip line** of the exterior wall sampled for paint lead measurements, a site approximately 12 inches from the structure; (3) in a **remote** location halfway between the drip line sample and the property line or fence of the next building (at least 5 feet from the building, but not more than 30 feet, and not taken from another property); and (4) in the **playground** area, if one existed.

Three soil subsamples were collected at each sampling location and combined in a plastic bag to yield one composite sample. Each subsample was collected by inserting a 1-1/8 inch diameter metal, tube-shaped corer into the ground and collecting the top 2-3 cm of soil. In order to consistently sample composite samples, the technician first selected the sampling location and collected the first of three subsamples. The second subsample was taken 20 inches to the right and the third, 20 inches to the left. Guidelines for soil sampling indicated that technicians were to avoid vegetation and runoff from potential lead sources (e.g., driveways) and select the sample from the same side of the dwelling as exterior XRF readings unless no soil was present. Prescribed soil sampling procedures were as follows:

- Team leader and technician always wore disposable gloves when taking soil samples.
- Gloves were changed after each soil sample was taken to avoid contamination.
- Technician inserted corer into ground approximately 10 centimeters. Sample consisted of only the top 2 to 3 centimeters.
- Technician put the three core subsamples in a plastic bag (one composite sample) and handed it to the team leader who double-bagged the sample, attaching the label to the inside bag.

- Technician cleaned the corer with a wet wipe after each sample was taken.
- Technician noted the location of the sample, and made remarks regarding exceptions to normal procedures on the Soil/Dust Sampling Log.
- If soil samples could not be taken as outlined in the procedures, the following guidelines were used: if there was soil within 25 feet of the sampled building, a soil sample was taken at that location and considered to be a remote sample.

2.2 Field Protocols - Implementation

A review of the National Survey data revealed that remote soil samples were taken in a variety of locations proximate to the dwelling units. Where remote samples could not be taken in a standardized manner, procedures directed technicians and interviewers to indicate deviations in the "comments" section of the Soil/Dust Log. Examples of comments indicated that remote soil samples were taken in flower beds, in gardens, near garbage bins, under a tree or bush, near a swing set or play area, or near a garage or parking lot. In cases where no comments were made it was assumed that guidelines indicated in the training materials were followed without incident.

As mentioned in section 1.2, both the drip-line and the remote soil samples were normally collected on the same, randomly selected side of the house as the XRF paint lead measurement was made. While this procedure was specified in the sampling protocols, its practice was not always possible. Table 2-1 summarizes cases for which the drip-line and the remote soil samples were collected on the same side of the dwelling unit, cases where they were not, and unrecorded (unspecified) cases. Assuming the unrecorded cases were sampled according to protocol, then summing the shaded diagonal row in Table 2-1 gives 182 (64%) drip-line and remote samples collected on the same side of the house. In some cases, especially in large cities, soil samples could not be collected. In these cases, technicians and interviewers were instructed to record why a sample could not be taken. Reported reasons included: soil was inaccessible; covered with concrete, stones or gravel; or that the ground was frozen. Note that the survey was collected on a tight schedule that involved collecting samples in winter. There were, therefore, a few unavoidable situations where samples could not be collected due to frozen ground.

The analytical methodology used for the survey was chosen on the basis of three criteria. The method should be (1) based on historical data from earlier lead studies (2) chosen on the basis of the estimated limit of detection needed to obtain useful data, and (3) based on an existing standardized methodology for each sample matrix. Following these guidelines, the laboratory selected the digestion procedures outlined in the standardized EPA SW-846 methodology for lead and the analysis procedures by inductively coupled plasma-atomic emission spectrometry (ICP-AES).

Table 2-1. Relative location of remote and drip line samples in the national survey

		Remote Wall Sampled						
		Wall 1	Wall 2	Wall 3	Wall 4	Unspecified	None Taken	Total
Drip-Line Wall Sampled	Wall 1	16	2	2	1	4	1	26
	Wall 2	0	3	8	2	8	1	22
	Wall 3	0	0	7	0	2	0	9
	Wall 4	2	0	2	10	5	0	19
	Unspecified	4	1	3	1	146	1	156
	None Taken	2	3	3	1	1	42	52
	Total	24	9	25	15	166	45	284

(1) Wall #1 faces the street of the dwelling units address, and continuing CLOCKWISE THE NEXT WALL IS #2, ETC.

(2) Assuming protocol was followed in unspecified cases, the number of cases where dripline and remote samples were taken off the same wall totals 182.

2.3 Laboratory Protocols

Midwest Research Institute (MRI) was responsible for laboratory analysis of the soil and dust samples collected in the National Survey. MRI and two Core Laboratories (one in Casper, Wyoming and one in Aurora, Colorado) analyzed the samples for lead. MRI's *Analysis of Soil and Dust Samples for Lead, Final Report, May 8, 1991* details their methodology and data quality procedures. The following was extracted from the MRI Report.

Internal and external checks, as part of MRI's Quality Assurance Project Plan (QAPjP), were used to track data performance. Checks included duplicate injections into the ICP-AES to measure instrument precision and the analysis of split samples to measure the variability from sample handling prior to analysis. Performance check samples (PCSs) of known lead concentration were also analyzed to measure accuracy of the analytical procedure. Identity of the PCS and split samples were unknown ("blinded") to the analytical laboratories in order to reflect a true representation of the routine analytical analysis.

MRI's final report stated that during the survey, a total of 1,053 soil samples were analyzed--139 duplicate injections; and 105 PCS and 105 split samples (10% of total). In general, the results for the PCS and the duplicate injections met the data quality objectives (DQOs) for both the Core Laboratories and MRI (+/- 30% for PCS and +/- 20% for duplicate injections). However, the data for the split samples did not meet the criterion of +/- 30%. Of the 105 split sample observations, 30% were outside the DQO. MRI's report states the probable reason for this variation is due to poor homogenization of the composite soil samples before splitting. Although the MRI report is not definitive on the subject, poor homogenization may have contributed to measurement variation for both the split samples and those which were not split.

3. Soil Data Results

This chapter provides descriptive statistics for the National Survey soil lead data. These statistics provide background information on the data used in the analyses discussed in Chapters 5 and 6. They are also used to assess the suitability of the data for other analyses that might be conducted in the future.

3.1 Recovery Bias Correction For The Soil Measurements

In order to measure the accuracy of the analytical soil lead analysis, three batches of soil were prepared with different lead concentrations at the beginning of the study to serve as ongoing method performance check samples (PCS). The soil PCS batches were prepared in the following manner: soil from a rural area was dried, mixed, and put through a 200-micron sieve. The soil was then split into three portions; the first one, referred to as the low-level spike, was actually not spiked (representing a background lead concentration of about 30 ppm). The second was spiked with 316.2 ppm of lead, and the third was spiked with 2,097 ppm of lead. These three batches of soil were referred to as the low-level spiked, mid-level spiked, and high-level spiked controls in the *Analysis of Soil and Dust Samples for Lead, Final Report.*[1]

Recoveries for the high- and mid-level spiked soils were calculated for this report using the following formula:

$$\frac{\text{Spiked sample measurement - Unspiked sample measurement}}{\text{Concentration change due to spike}}$$

The average lead measurements in the spiked and unspiked samples were calculated from MRI's analysis report.[2]

Table 3-1 shows the recoveries for the soil control samples. Recoveries are slightly, although significantly, below 100 percent and depend on both the lab performing the measurement and the spiking concentration. Recovery was higher for MRI than for the Casper lab and recovery decreased as the lead level in the PCS samples increased. Because the soil lead concentrations were generally less than 300 ppm found in the mid-level spike and rarely as high as 2100 ppm found in the high-level spike, the measured recovery for the mid-level spike was used to correct the soil lead measurements for recovery.

---

[1]Appendix C of: *Analysis of Soil and Dust Samples for Lead (Pb)*, MRI, May 8, 1991. Note that the soil portion referred to as the low-level spike was not actually spiked with lead. The spiking levels were reported by MRI in a phone call to Westat.

[2]MRI reported the ratio of the average lead measurement to the measurement determined at MRI prior to the sample collection. This ratio was labeled "percent recovery" however is not actually a measurement of recovery. The average lead measurement was calculated from this ratio.

Table 3-1. Percent recovery for soil control samples

Lab	Sample Size	Averaged measured concentration ppm	Spiked concentration ppm	Recovery (percent) with 95% conf. interval	Comments
Casper	21	26.8 ± 2.6			Unspiked samples for recovery calculation
	18	300 ± 9	316.2	86.4% ± 3.0%	
	23	1720 ± 170	2097	80.7% ± 8.1%	Includes 1 low outlier
MRI	14	33.2 ± 1.7			Unspiked samples for recovery calculation
	15	336 ± 12	316.2	95.8% ± 3.8%	
	14	1985 ± 109	2097	93.1% ± 5.2%	

Based on ICP-AES analysis of soil control samples. Source: Analysis of Soil and Dust Samples for Lead (Pb), MRI, May 8, 1991.

Note: ± values are 95% confidence intervals for the respective parameters.

All the results presented in this report are based on the recovery corrected lead concentrations. Use of a recovery correction provides measurements of lead concentration that are relatively unaffected by the choice of laboratory or laboratory measurement technique.

3.2 Description Of The Soil Data

This section introduces the first of a two-part statistical description of the soil lead data from the National Survey. Univariate findings concerning the distribution of the data, outliers, censoring, and basic descriptive statistics are presented. Section 3.3 then describes the relationships between the soil lead values at each of the three sampling locations. Neither analysis reveals significant statistical problems that would compromise the utility of the data for other analyses.

Additional descriptive analyses of the soil lead data may be found in the two reports, *Comprehensive and Workable Plan for the Abatement of Lead-Based Paint in Privately Owned Housing: Report to Congress,* prepared by HUD, and *Report On The National Survey Of Lead-Based Paint In Housing,* prepared by Westat for EPA.

3.2.1 Outliers

Outliers are observations that are unusual compared to other comparable observations. Outliers may be valid observations on samples collected at unusual sites. These outliers can point out situations that may be of interest. Alternatively, outliers may result from errors in the sample collection, analysis, or data recording procedures and thus not represent the true concentration at the site. These outliers are of no interest and should not be used in any analysis. Unfortunately, these two types of outliers cannot be distinguished based on the data. In addition, sometimes incorrect observations cannot be distinguished from correct observations and do not appear as outliers. As a result, outlier identification procedures cannot identify observations that are incorrect. They can only identify observations that should be looked into more carefully and that may affect subsequent statistical analysis.

The general procedure for identifying outliers involves the following activities:

(1) Specifying a model for the observations, including specifying the distribution of the residuals (the difference between the observations and the model prediction).

(2) Fitting the model to the data, calculating the residuals.

(3) Testing to determine if the most extreme residuals are unusual given the assumed distribution of the residuals.

These steps may be repeated using several assumed models for the data.

In order to identify outliers, it is necessary to specify the model that describes the data and the residuals. This model describes the values against which an observation is compared to determine if it is unusual. For example, one drip line soil measurement may not be unusual compared to all other drip line soil measurements, but may be unusual compared to all other drip line measurements for dwelling units built in the same decade. Similarly, although the drip line and remote measurements may not be unusual when compared to other drip line and remote measurements, the pair of observations may be unusual when compared to pairs of observations from all dwelling units.

For the outlier tests used in this report, the residuals were transformed to variables with an approximately normal distribution. The specific transformation is discussed below. Then, the extreme studentized deviate was used to test for outliers (see *Statistical Methods*, Snedecor and Cochran, 1980, 7th ed., page 280). This test assumes that the residuals are independent and have a normal distribution. For a one-sided test of the largest residual, an observation is assumed to be an outlier if the probability of observing a maximum residual as large or larger than the observed maximum is less than one percent. Because tables of the extreme studentized deviate for large sample sizes were not readily available, critical values were calculated using a normal approximation. For a set of 250 measurements (roughly the number of dwelling units with soil samples) the maximum residual is assumed to be an outlier if it is more than 3.94 standard deviations from the mean. For a two-sided one percent test, looking for either large or small outliers, the critical value is 4.11 standard deviations. Because 1) the residuals are not completely independent, 2) the distribution of the residuals is only approximately normal, and 3) multiple tests, rather than one, were used to identify outliers, the probability of incorrectly identifying the extreme observation as an outlier is only approximately equal to the nominal level of one percent.

The following three models were used to identify possible outliers:

- Assume the entrance, drip line, and remote measurements have a log normal distribution. Fit a mean to the log-transformed measurements and assume that the residuals have a normal distribution.

- Assume the log-transformed entrance, drip line, and remote measurements have a multivariate normal distribution. Calculate the Mahalanobis distance (see discussion below) for each dwelling unit. Assume that the cube root of the Mahalanobis distance has a normal distribution.

- Assume the residuals from a two-way analysis of variance model on the log-transformed entrance, drip line, and remote measurements, with factors for dwelling unit age and county, have a multivariate normal distribution. Calculate the Mahalanobis distance for each dwelling unit. Assume that the cube root of the Mahalanobis distance has a normal distribution.

A scatter plot of two variables typically forms a cluster of points. With three variables, the scatter plot would be a cloud of points in space. For identifying outliers using several variables, the Mahalanobis distance measures the distance of each observation from the center of the cloud of points (when plotting in 2, 3, or more dimensions). The distance measure is a function of the standard deviation and correlation of each of the

variables. Section 4.3.1 provides an example of the use of similar procedures on the dust data, including a graphical example of the Mahalanobis distance.

Although three dwelling units had somewhat unusual measurements based on visual inspection, no dwelling unit or individual measurement was classified as an outlier using the extreme studentized deviate test. We checked the records for the dwelling unit with the largest Mahalanobis distance. The only indication of an unusual situation was that the soil samples were taken from frozen ground. Because frozen ground was encountered in other dwelling units in the same county and those dwelling units did not have unusual soil lead measurements, there was no rationale for questioning the data.

Based on this review of the data, no observations were identified as outliers and the most extreme observations showed no identifiable problems. Therefore, the subsequent analyses use all of the soil measurements and no reason was found to recommend excluding any soil measurements from other possible analyses.

3.2.2 Distribution Of The Soil Lead Measurements

The soil measurements have a skewed distribution, with many low lead concentration measurements and few relatively high measurements. The distribution can be roughly described by a log normal distribution. For the log normal distribution, the log-transformed data (that is the natural log of the measured lead concentrations) has a normal or bell shaped distribution. Figures 3-1, 3-2, and 3-3 show histograms of the lead concentrations for the entry way, drip line, and remote soil samples respectively, using a log scale. The histograms suggest that the data are slightly skewed to the right even after using the log transformation.

3.2.3 Censoring

A total of 24 soil lead concentrations were reported below the detection limit. This represents three percent of the soil lead measurements. For the analysis of the soil lead data, a common practice was followed of replacing the measurements below the detection limit with one-half of the detection limit. In all cases, half of the detection limit was less than or equal to 5 ppm. Because the number of measurements below the detection limit is small and the use of one-half of the detection limit appears to be consistent with the distribution of all the measurements (see Figures 3-1, 3-2, and 3-3), the handling of measurements below the detection limit is expected to have no significant effect on the statistical analysis results.

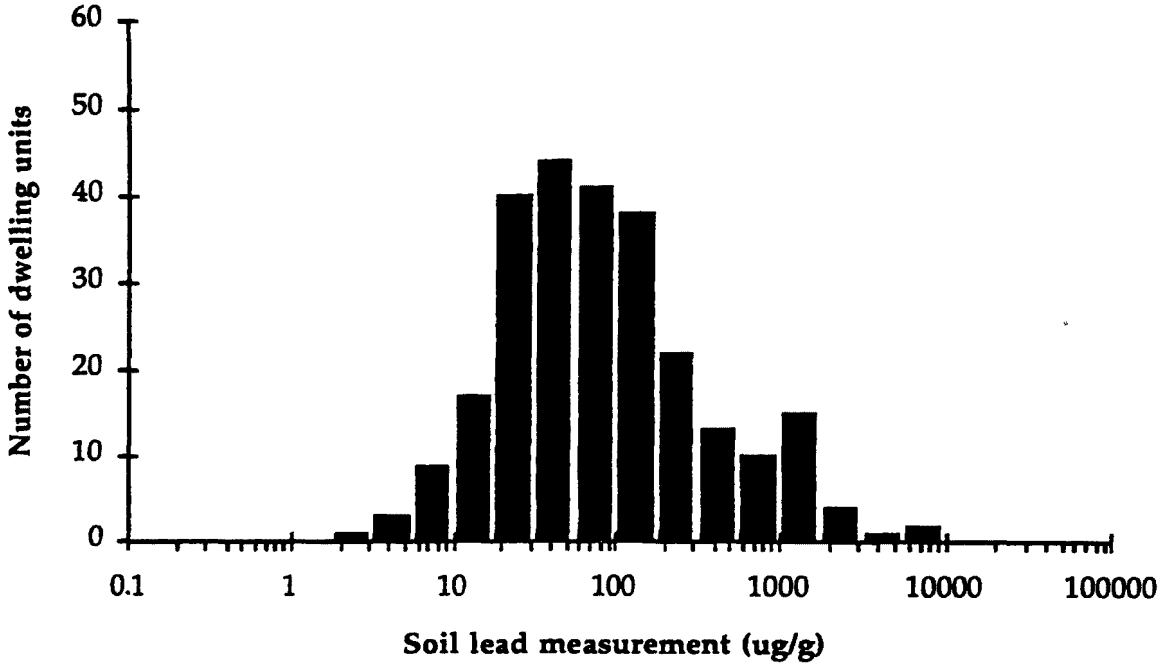


Figure 3-1 Distribution of the lead measurements in soil samples collected outside the dwelling unit entrance

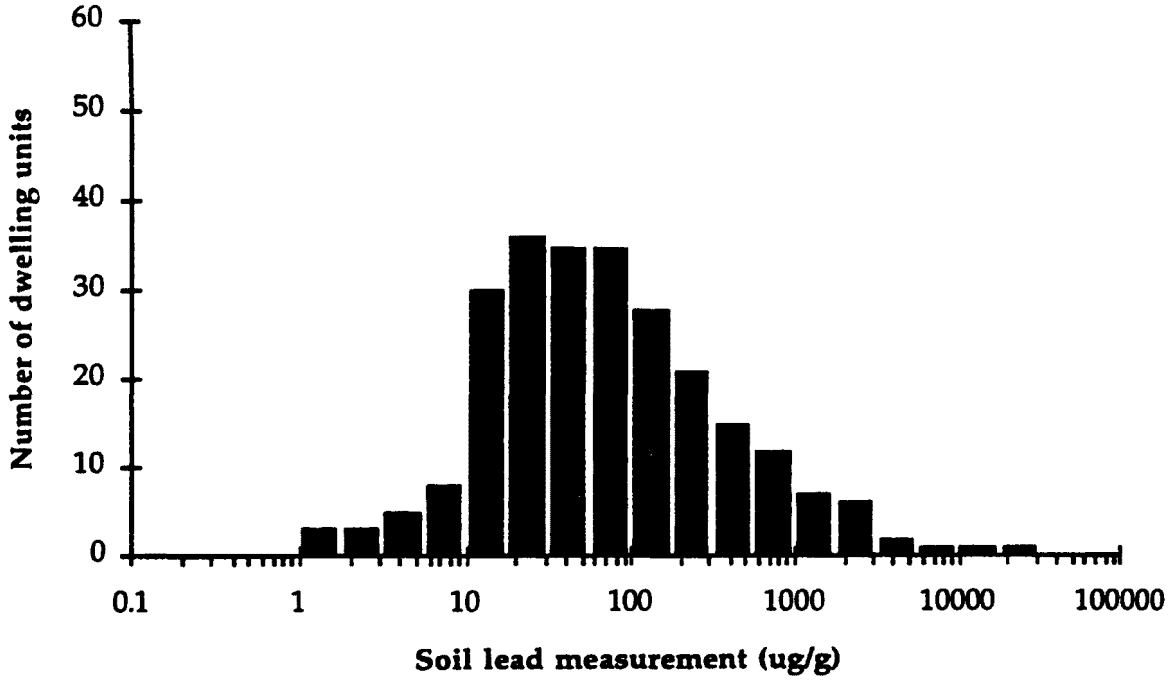


Figure 3-2 Distribution of the lead measurements in soil samples collected at the drip line

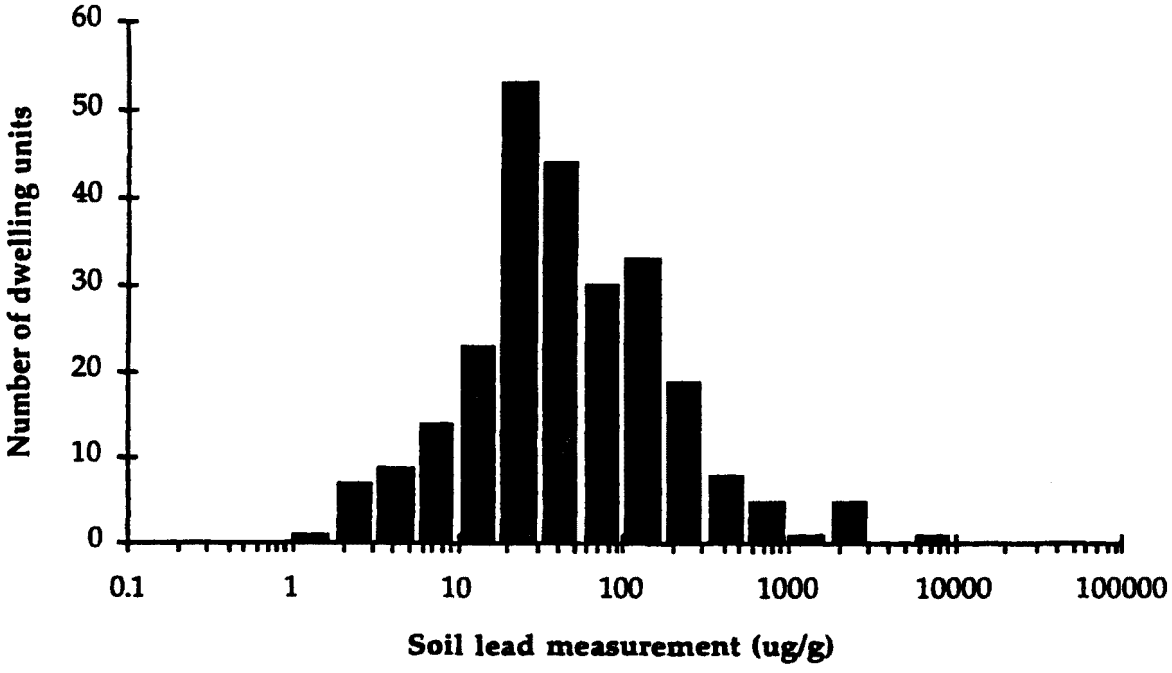


Figure 3-3 Distribution of the lead measurements in soil samples collected at remote locations away from the dwelling unit

3.2.4 Means Standard Deviations, And Descriptive Statistics

Table 3-2 summarizes the sample mean, standard deviation, coefficient of variation, selected percentiles, geometric mean, and standard deviation of the log-transformed measurements for the entrance, drip line, remote, and playground soil measurements. The mean, standard deviation, and coefficient of variation are based on the soil lead concentration measurements. The coefficient of variation is the ratio of the standard deviation to the mean of the data and describes the spread of the measurements relative to the average. The coefficient of variation is useful for describing data that have skewed distributions and are always greater than or equal to zero. The geometric mean is the mean of the log-transformed measurements expressed in the untransformed scale. If the data has a log normal distribution, the geometric mean approximates the median of the distribution.

The geometric mean soil lead concentrations for the entrance, drip line, and remote locations are 83, 72, and 47 ppm respectively. However, soil lead concentrations at individual sites can vary considerably around these mean levels. Because of the small number of playground soil samples (6), the playground samples were not included in the analysis of soil measurements.

Because the survey dwelling units were selected using a multistage probability sample, inferring the summary statistics for dwelling units nationally requires using weighted statistics to account for differences in sampling rates and the effects of the complex sample design.[1] The weights depend on the probabilities of selecting individual homes in the survey. The unweighted results in Table 3-2 describe soil measurements collected from the specific homes sampled in this survey. Table 3-3 contains descriptive statistics for the **weighted** soil lead measurements, which represent statistics for dwelling units nationally.

Results were also explored by age of the dwelling units. Preliminary analysis suggested that soil lead concentrations are greater for older dwelling units. The survey data supports this suggestion and can be seen in Figures 3-4, 3-5 and 3-6 -- histograms of the weighted soil lead concentrations, broken down by dwelling unit construction year, for the entrance, drip-line, and remote soil samples, respectively. Tables 3-4 and 3-5 summarize these results by dwelling unit age. These tables present weighted arithmetic and geometric means with 95 percent confidence intervals.

Weighting is required for computing unbiased estimates of parameters of interest, e.g., the estimated national geometric mean entryway soil lead concentration is 85 ppm. On the other hand, weighting is generally not necessary for conducting relational analyses like the correlations and regressions in this report. It is complicated to correctly weight these analyses and the results often do not warrant the effort. Therefore, the correlations and regressions in this report are unweighted.

---

[1] See *Comprehensive and Workable Plan for the Abatement of Lead-Based Paint in Privately-Owned Housing: A Report to Congress* for a description of the sample design.

Table 3-2 Descriptive statistics for the lead measurements in soil samples (unweighted)

Set of data	Entrance samples	Drip line samples	Remote samples
Number of measurements	260	249	253
Arithmetic mean (ppm) (95% confidence interval)	295 (203 to 387)	415 (192 to 638)	170 (99 to 240)
Standard deviation (ppm)	753	1790	570
Coefficient of variation	2.55	4.31	3.35
Percentiles (ppm) maximum upper quartile median lower quartile minimum	 6,829 199 64.8 30.5 2.84	 22,974 199 60.2 23.1 1.16	 6,951 119 42.8 19.3 1.45
Geometric mean (ppm) (95% confidence interval)	83 (70 to 100)	72 (58 to 89)	47 (40 to 56)
Mean of the log-transformed measurements	4.42	4.27	3.85
Standard deviation of the log-transformed measurements	1.47	1.68	1.42

Table 3-3 Descriptive statistics for the lead measurements in soil samples (weighted)

Set of data	Entrance samples	Drip line samples	Remote samples
Number of measurements	260	249	253
Arithmetic mean (ppm)	327	448	205
Percentiles (ppm)			
maximum	6,829	22,974	6,951
upper quartile	225	230	119
median	64.6	56.3	39.9
lower quartile	28.4	21.2	18.5
minimum	2.84	1.16	1.45
Geometric mean (ppm)	85	74	46
Mean of the log-transformed measurements	4.44	4.31	3.84

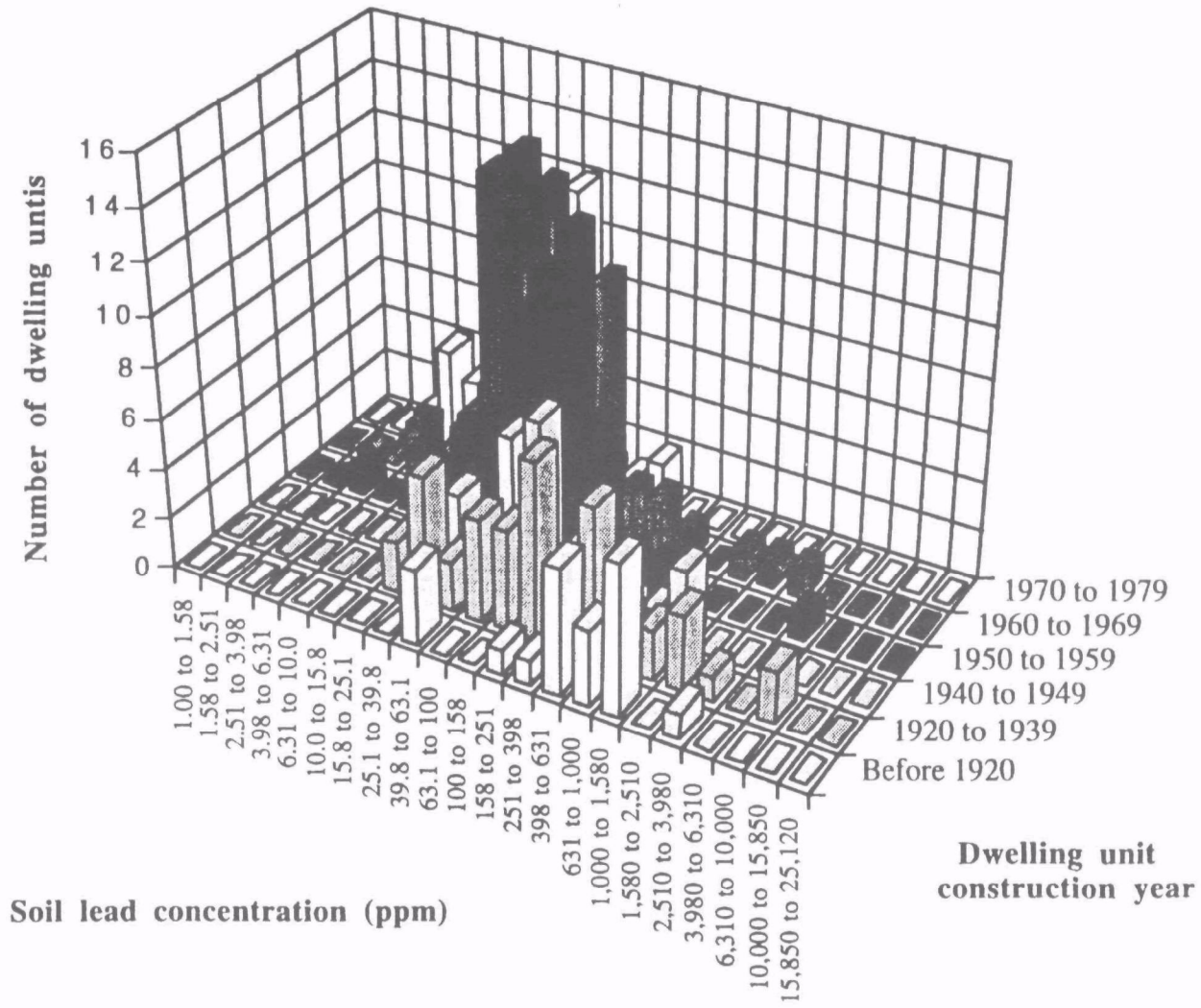


Figure 3-4 Histogram of entrance soil lead concentrations by dwelling unit construction year.

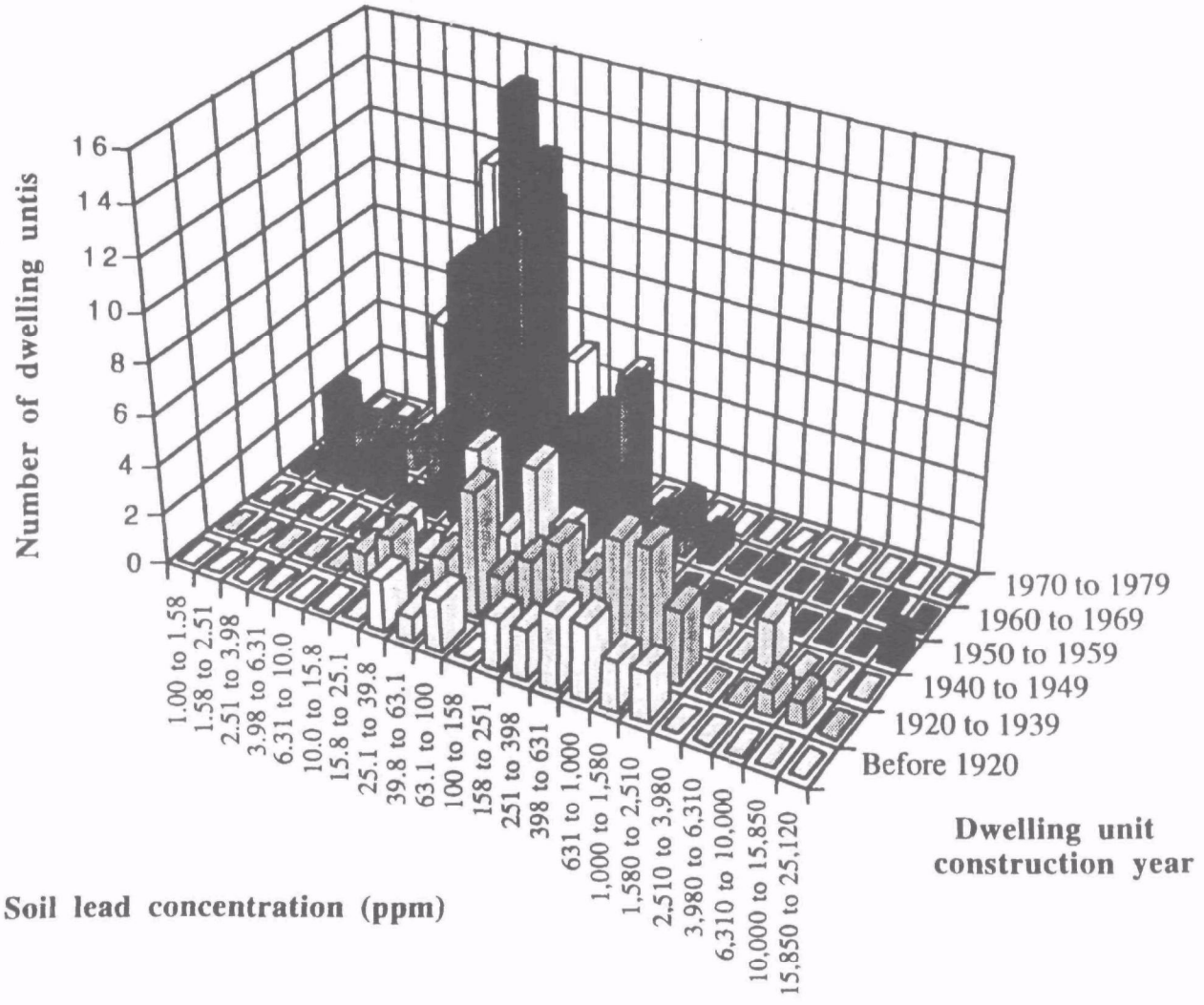


Figure 3-5 Histogram of drip-line soil lead concentrations by dwelling unit construction year.

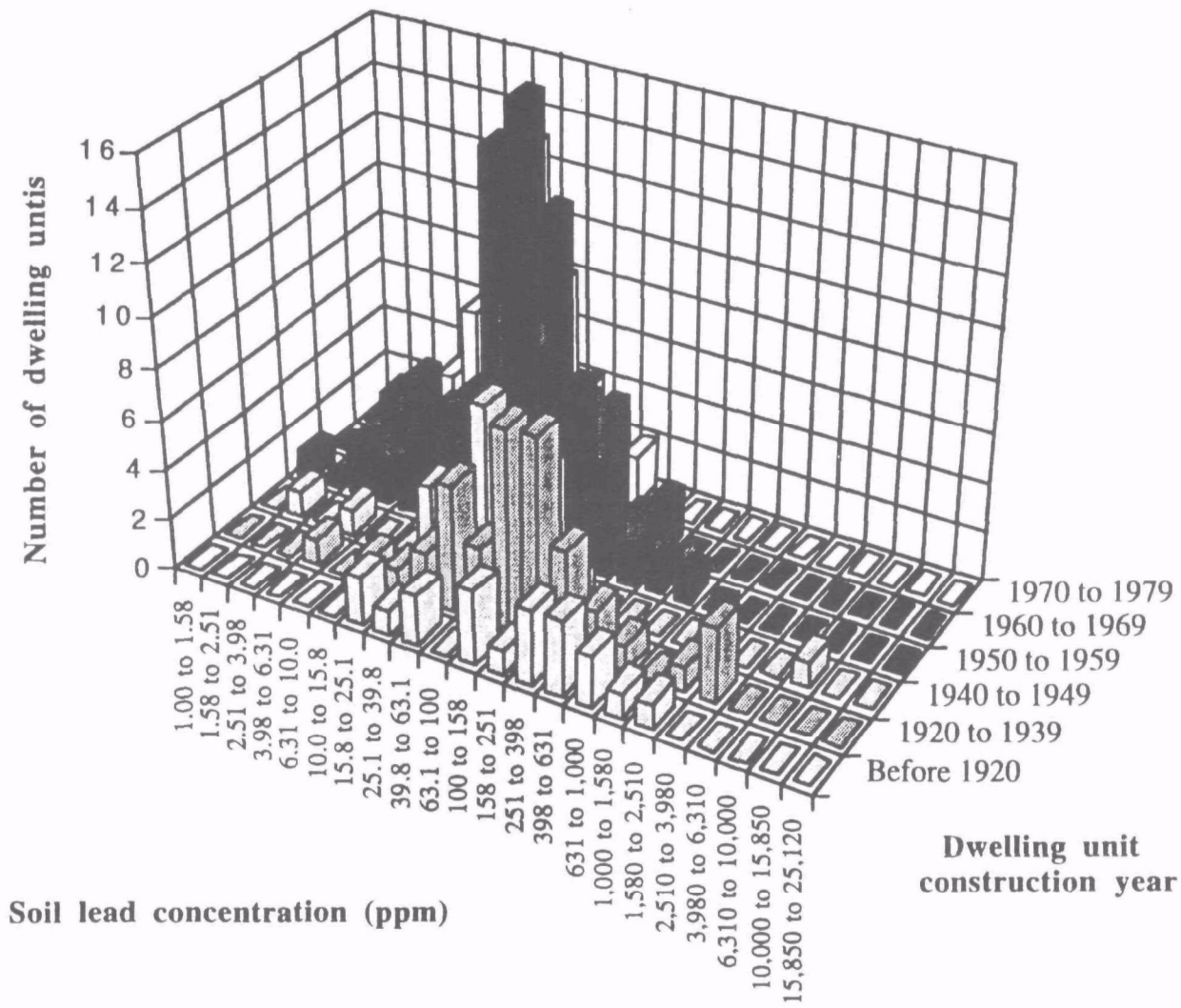


Figure 3-6 Histogram of remote soil lead concentrations by dwelling unit construction year.

Table 3-4 Arithmetic mean soil lead concentrations by dwelling unit age, with 95% confidence intervals (weighted)

Construction Year	Entrance samples	Drip line samples	Remote samples
Before 1920	669 (424 to 915)	587 (333 to 842)	426 (179 to 674)
1920 to 1939	1015 (347 to 1684)	1488 (426 to 2549)	552 (193 to 911)
1940 to 1949	272 (89 to 455)	447 (24 to 870)	532 (0 to 1371)
1950 to 1959	168 (0 to 346)	508 (0 to 1514)	87 (45 to 129)
1960 to 1969	131 (35 to 226)	68 (35 to 102)	44 (29 to 59)
1970 to 1979	40 (25 to 55)	36 (22 to 50)	27 (19 to 36)
All Dwelling Units	327 (195 to 458)	448 (183 to 714)	205 (102 to 308)

Confidence intervals assume variance inflation factor of 1.45.

Table 3-5 Geometric mean soil lead concentrations by dwelling unit age, with 95% confidence intervals (weighted)

Construction Year	Entrance samples	Drip line samples	Remote samples
Before 1920	503 (302 to 837)	383 (200 to 732)	248 (125 to 492)
1920 to 1939	329 (189 to 572)	534 (289 to 985)	159 (85 to 297)
1940 to 1949	135 (80 to 229)	151 (77 to 295)	67 (31 to 145)
1950 to 1959	74 (53 to 105)	70 (43 to 112)	44 (29 to 67)
1960 to 1969	49 (35 to 68)	31 (21 to 44)	25 (18 to 34)
1970 to 1979	27 (20 to 37)	22 (15 to 31)	20 (15 to 27)
All Dwelling Units	85 (68 to 106)	74 (57 to 97)	46 (37 to 58)

Confidence intervals assume variance inflation factor of 1.45.

3.3 Interrelationships In The Soil Data

3.3.1 Correlations Between Locations

The measurements at the three locations, entrance, drip-line, and remote, are all highly correlated with each other. The correlations are shown in Table 3-6. Figures 3-4, 3-5, and 3-6 show scatter plots of the measurements. The points in each plot tend to scatter around a line, consistent with the highly significant correlations between the measurements at different locations.

3.3.2 Differences Among Sample Locations

For most dwelling units there are soil lead measurements at all three locations, the entrance, drip line, and remote location. The paired differences between the log-transformed measurements were used to determine if the geometric means at different locations were statistically significant. Based on the log-transformed measurements, the average soil lead concentration at the remote location is lower than that at either the entrance or drip line location. These differences are statistically significant at the 0.001 level. The differences between the entrance and drip line measurements are not statistically significant.

The remote samples have lower lead concentrations than the entrance and drip line samples. This might be explained by either a contribution of the paint on the dwelling unit to the nearby soil or possibly lead from airborne sources washing off the roof and walls of the dwelling unit and concentrating near the drip line, or a combination of these.

3.3.3 Measurement Variation

Each reported soil lead measurement has associated with it measurement error (also called measurement variation). As a result of this measurement error, the reported lead measurement will be different from the actual lead concentration in the soil sample or the average lead concentration in the vicinity of the soil sample. Although it is impossible to know the difference between the measurement and the actual concentration for any one sample, it is possible to estimate the average magnitude of the measurement error. Both the variance and the standard deviation measure the magnitude of the measurement variation. The standard deviation has the same units as the data and is the square root of the variance.

For this discussion it is assumed that the variable of interest is the average soil lead concentration in the vicinity of the entrance, the vicinity of the drip line, and in remote locations.

Table 3-6 Correlations between log-transformed soil lead measurements from different locations around the same dwelling unit

	Soil lead measurements (ppm)		
	Ext. entrance	Dripline	Remote
Soil lead entrance ppm		0.7148 0.0001 246	0.6090 0.0001 247
Soil lead dripline ppm	0.7148 0.0001 246		0.6780 0.0001 243
Soil lead entrance ppm	0.6090 0.0001 247	0.6780 0.0001 243	

Note: In each set of table entries, the top number is the correlation coefficient, the middle is the probability that a sample correlation this far from zero might occur by chance if there were actually no correlation in the underlying population, and the bottom number is the number of paired measurements used to calculate the correlation.

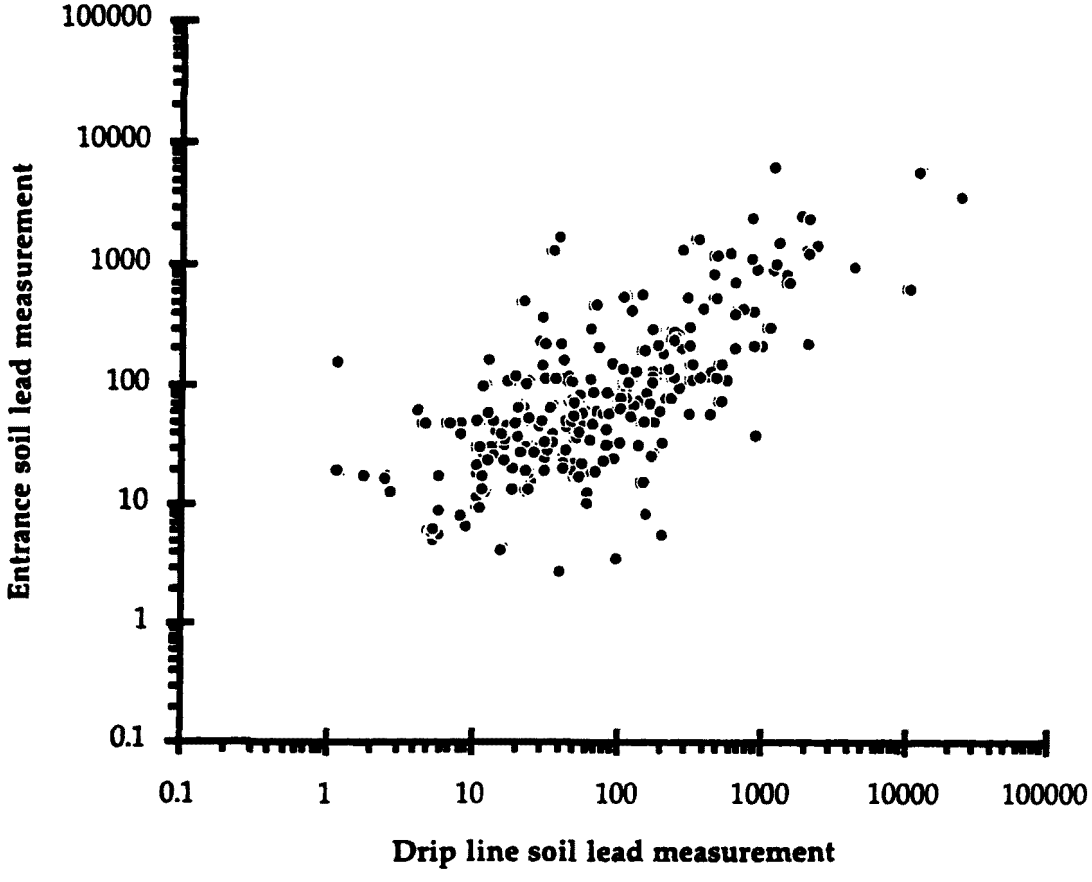


Figure 3-7 Plot of soil lead measurements at the entrance location versus at the drip line location

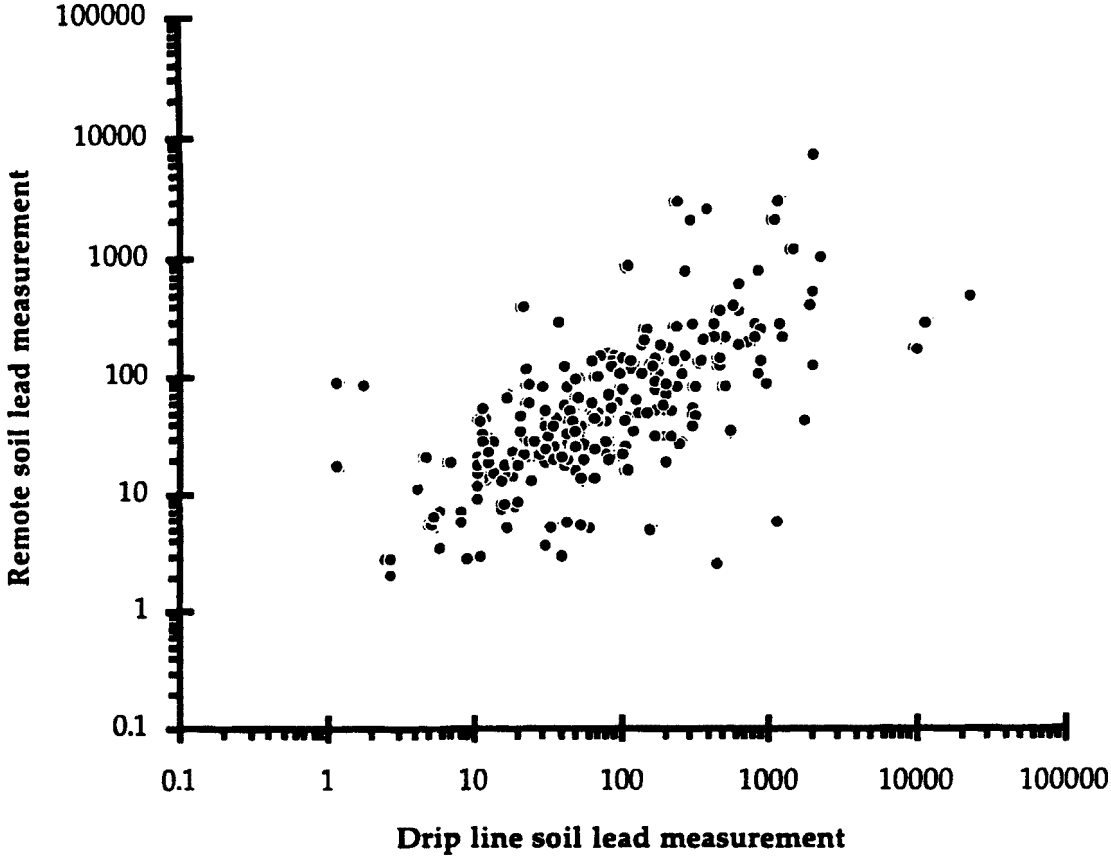


Figure 3-8 Plot of soil lead measurements at the remote location versus at the drip line location

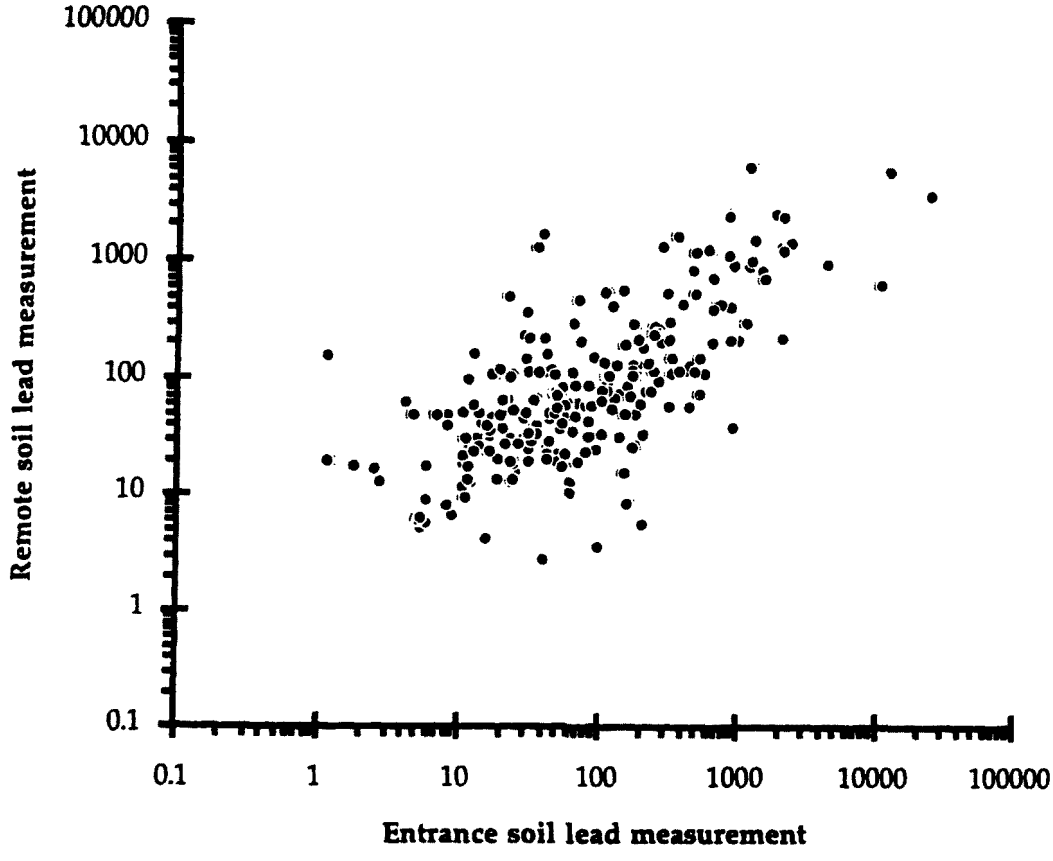


Figure 3-9 Plot of soil lead measurements at the remote location versus at the entrance location

Each soil sample was a composite of three soil cores collected roughly 20 inches apart. Each core was 2 to 3 centimeters deep and 1-1/8 inches in diameter. Because this sample covers 3 square inches of soil, its lead measurement only approximates the average concentration over the larger area of interest. Because the soil lead concentration is likely to vary across the area of interest, the actual lead concentration in the sample may be different from the average lead concentration across the area of interest. This difference contributes to the measurement error. When the soil sample gets to the lab, only a portion of the sample is actually analyzed. Because different portions of the sample will have slightly different lead concentrations and because of uncontrolled variation in the sample preparation and final measurement step, the sample preparation and measurement procedures will also contribute to the measurement error.

The variance of the measurement error can be approximated by the following model:

Variance of the log-transformed measurement error = $\sigma^2 = \sigma^2_{field} + \sigma^2_{lab}$.

where:

σ^2_{field} = the variance of the difference between the average concentration over the area of interest and the lead concentration in the field sample sent to the lab, after log transformation; and

σ^2_{lab} = the variance of the difference between the lead concentration in the field sample sent to the lab and the laboratory lead measurement, after log transformation.

Several of the soil samples were split into two portions. Measurements were made on each of the two portions. From these measurements we can estimate σ^2_{lab}. The soil samples were analyzed by two different laboratories. Because the lab may contribute different amount of variation to the measurement, the variances are summarized separately for each lab. For the Casper lab, which analyzed most of the soil samples, the estimated variance contributed by the lab to the log-transformed measurements is 0.060. For the MRI lab, the estimated variance contributed by the lab to the log-transformed measurements is 0.132. For both labs the variance in the log-transformed measurements appears to be constant, independent of the lead concentration.

The contribution of the field sampling to the variance of the log-transformed measurements (σ^2_{field})is more difficult to estimate. Two factors contribute to the difficulty in estimating the sampling component of variance: (1) there are no measurements from which to estimate the variance directly, and (2) the variance will be a function of the size of the area over which the soil lead concentration is averaged.

Presumably, lead concentrations at locations that are close together are similar. As the area over which the samples are taken increases in size, the variability of the lead concentration among randomly located samples increases. We can use the ratio of the measurements at two different locations to show the effect of distance between the samples on the variability of the measurements. We must assume that (1) σ^2_{field} is the same for the entrance, drip line, and remote locations (this assumption is often made for this type of data) and (2) that the ratio of the measurements from different locations is

constant, independent of other factors such as the presence of lead-based paint. The assumption of a constant ratio is consistent with scatter plots in Section 5.4, which show a linear relationship between the soil measurements at different locations and for which the slope of the linear relationship is approximately 1.0. If the assumption that other factors do not affect the difference between soil measurements is incorrect, the estimates of σ^2_{field} will tend to be larger than the actual value.

Figure 3-10 shows the variance of the ratio of log-transformed lead measurements as a function of distance. While the exact distance between the soil samples cannot be determined from the survey data, the measurements can be grouped as (1) on the same side of the dwelling unit, (2) on adjacent sides, or (3) on opposite sides. These groups of data correspond roughly to increasing distance between the sample locations. In dwelling units where the information is incomplete on which side of the house the samples were taken, we assumed the survey staff followed the standard procedures by which the entrance was designated side number 1 and the remote sample and drip line samples were taken on the same side of the house as the sampled exterior surface. As a result of the incomplete data and these assumptions, the analysis of the variances of the ratio between different sampled surfaces is approximate.

The results in Figure 3-10 suggest that the sample-to-sample variance does not depend greatly on the distance between the samples when the distance is greater than the typical distance between drip line and entry way samples on the same side of the dwelling unit. For shorter distances, the sample-to-sample variance decreases down to a small variance between side-by-side samples (represented by the spilt samples).

The variance of the difference between log-transformed measurements is the sum of the individual variances. Therefore the variance of one soil measurement for estimating the average soil concentration across as area (σ^2) is half of that shown in Figure 3-10. Because the variance depends on the size of the area of interest and because the values shown are approximate, for subsequent analyses the following assumption was made: the standard deviation of the log-transformed soil lead measurements around the average log-transformed soil lead concentration (σ) is 0.50 (the variance is therefore 0.707, and the variance of the difference between values is 1.41). Assuming the soil lead measurement errors have a log normal distribution, we would expect 95 percent of the soil lead concentrations to be within a factor of 2.7 of the true average lead concentration in the area of interest. For example, if a soil lead measurement is 47 ppm, the true lead concentration in the area of interest around the soil sample is most likely between 17 and 127 ppm (47/2.7 and 47*2.7 respectively).

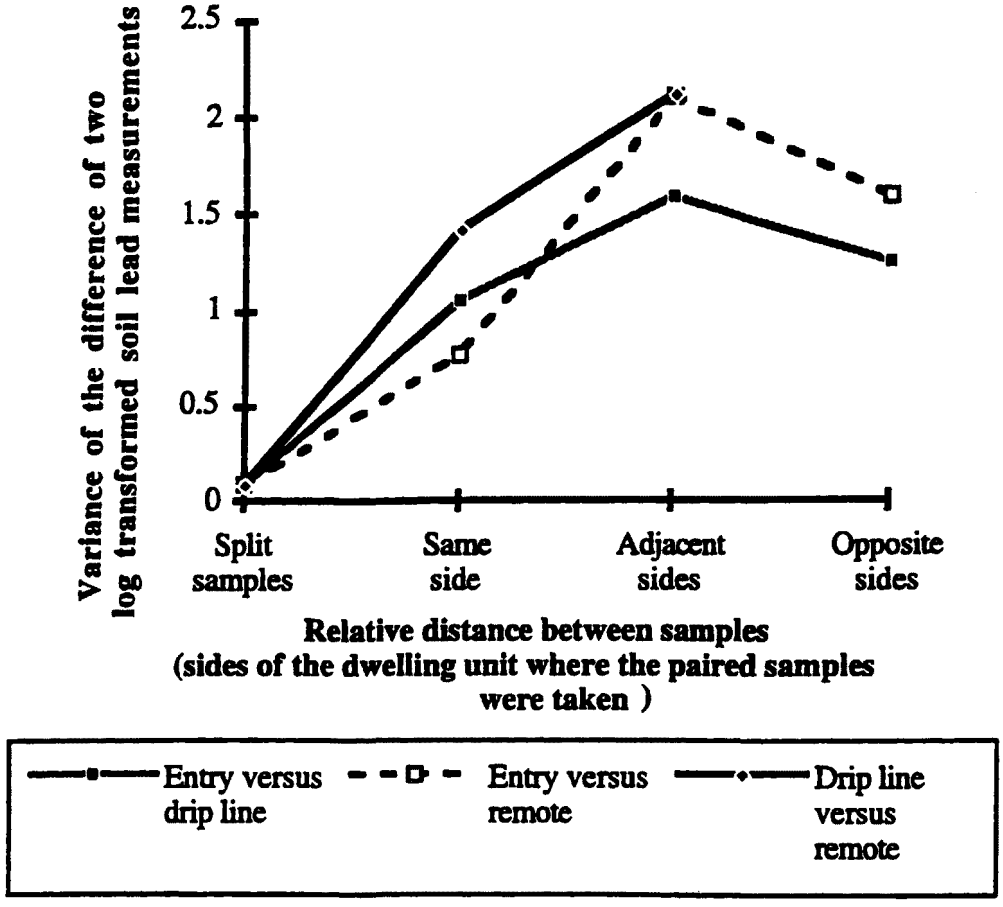


Figure 3-10 Approximate sample-to-sample variance as a function of relative distance between the samples

3.4 Limitations In The Data

The soil data have no serious statistical problems. Neither censoring nor outliers appear to significantly affect the data. After correcting for recovery, the measurements can be assumed to be unbiased. Although the measurement error might be considered to be large (i.e., within a factor of 2.7), the measurement error is small compared to the differences in soil lead concentrations between homes of different ages. In spite of the measurement error, significant correlations and differences between locations can be identified.

Interpretation of the results should reflect how the samples were taken. Samples were not taken where there was no soil near the sample location, for example if there were only concrete in the vicinity of the entrance. Thus the samples only represent dwelling units that are not surrounded by cement. The definition of the sample locations (entrance, drip line, and remote) was interpreted in the field to apply to each sampling situation. The survey records typically do not have details on exactly where the samples were actually taken. Therefore the classifications of entrance, drip line, and remote are rough descriptive terms to describe the sample location rather than indications of a carefully defined location.

Since the selection of dwelling units was based on a multistage probability sample, inference from the sample to the population of dwelling units nationally should be based on a weighted analysis reflecting the sample selection weights. For the purpose of estimating the relationship between measurements or variables from the same dwelling unit, use of the sample weights generally is not justified.

With the qualifications as to the locations at which the samples were taken and the sample of homes involved, the data show no patterns that might restrict the use of the data in subsequent statistical analyses.

4. Sampling And Laboratory Procedures For Dust Data

This chapter presents a brief description of the protocols used in the National Survey to collect the dust samples and analyze them for their lead content. The description is focused on establishing the utility and appropriateness of the data for the dust lead analyses of Chapter 6. Also presented are descriptive dust lead statistics.

Additional information and data can be found in the three reports, *Comprehensive and Workable Plan for the Abatement of Lead-Based Paint in Privately Owned Housing: Report to Congress*, prepared by HUD, *Report On The National Survey Of Lead-Based Paint In Housing*, prepared by Westat for EPA, and *Analysis of Soil and Dust Samples for Lead (Pb)*, prepared by MRI for EPA.

4.1 Dust Sample Collection And Analysis Procedures

Currently, there is no standard dust sampling technique to sample lead in household dust. However, for the National Survey, MRI experimented with previously described techniques found in the literature and experimented with their own designs. The method developed to collect dust in this study was designed to measure dust loadings (micrograms of lead per square foot of sampled surface), not dust lead concentrations (micrograms of lead per gram of dust). However, the equipment did allow making lead concentration estimates. The major component of the sampler chosen was a Gast rotary-vane vacuum pump that ran on standard 110 volt wall current. The sample train consisted of the pump, thick-walled 3/8" Tygon tubing connected to the vacuum side of the pump, and a 37-mm mixed cellulose ester membrane filter cassette (0.8 μm pore size) connected to the other end of the tubing. A specially designed angle cut Teflon nozzle (4" long x 2" wide) was developed to be inserted over the filter cassette (open-faced cassette and sealed with o-rings) and the sampling flow rate was set in the lab to approximately 16 liters per minute. The nozzle attachment was changed with each new sample to prevent cross-contamination.

Dust samples were collected at seven locations within the interior of each sampled housing unit: the floor of the dry room, a window sill in the dry room, a window well in the dry room, the floor of the wet room, a window sill in the wet room, a window well in the wet room, and the floor just inside the main entrance to the housing unit. See Section 1.2 of this report for definitions of window sills and window wells. Windowless rooms were sometimes encountered; only floor dust samples were taken in these rooms. Additional dust samples were collected in the common areas of multifamily housing units: the floor of the common hall just outside the sampled housing unit, the floor of the common hall just inside main entrance to the building, the floor of a randomly sampled common room (e.g., mail room, laundry room, lobby, etc.), a window sill in the same common room, and a window well in the same common room. However, because the number of dust samples collected in common areas was small, they are not included in the statistical analysis for this report.

All floor samples were collected in the following fashion. The field interviewer and technician both donned disposable plastic gloves. A 12" by 12" template was placed on the floor at the location in the room where the largest quantity of dust was expected to be found. Carpeted areas were preferred. If no rug was present, the dust sample was taken from the hard floor. The area within the template was vacuumed in overlapping passes, first left to right over the entire area and then front to back. The procedure was repeated until a four-square foot area had been vacuumed. Between templates, the vacuum nozzle was turned upward and left on to prevent dust from falling out of the cassette. After sampling, the field technicians were instructed to visually inspect the filter cassette to determine if enough dust for the laboratory analyses had been vacuumed. The technician determined this by comparing the filter cassette with photographs of filter cassettes containing varying sufficient and insufficient amounts of dust. If not enough dust had been vacuumed, the technician was instructed to vacuum additional template areas until enough dust was in the filter. In most cases, vacuuming four square feet yielded sufficient dust. In all cases, the actual floor area vacuumed was recorded. The gloves were changed after each dust sample was taken.

Window sill and window well dust samples were collected in a similar manner, except that the template could not be used in these locations. The field staff simply measured and recorded onto data sheets the length and width of the areas vacuumed from windows.

After the vacuuming was completed, the cassette ports were plugged and the cassettes were labeled and double-bagged in plastic bags for shipment to the laboratory.

Prior to data collection in the National Survey, it was felt that a more sensitive analytical procedure than ICP-AES would be needed to measure lead in dust because the dust collection methodology could not guarantee that enough dust would be consistently collected for the ICP-AES analysis. Therefore, graphite furnace atomic absorption (GFAA) spectroscopy was used to analyze dust. In this technique, sample digestate is atomized in a graphite furnace at high temperatures. This procedure is the most sensitive of the analytical techniques specific for lead.

The field technician collected the dust sample from areas where dust was available, particularly for the floor samples. As a result, the dust samples do not represent a random selection of dust from floors of the sampled rooms and the estimates of dust and lead loadings will provide biased estimates of average dust and lead loadings across the floor. The dust loadings reflect the locations where the samples was taken, which were locations likely to have greater dust loadings than other unsampled locations. Because the lead loading tends to increase as the dust loading increases, the lead loading estimates will also tend to overestimate the average lead concentrations across the floor.

4.2 Recovery Bias Correction For The Dust Data

Unlike the soil performance check samples (PCS), no standard dust samples were available for use as PCS. Therefore, the soil PCS were used as "dust" to verify the accuracy of the dust analysis during the study. The spike levels in these "dust" samples were therefore the same as in the soil PCS described in Section 3.1 of this report. Because the soil PCS visually look different from dust samples, they could not be "blinded" to the laboratory conducting the analysis. Thus, recovery results may reflect extra care by the laboratory in analyzing "dust" PCS, causing recoveries to be higher than expected.

Table 4-1 shows the recoveries for the dust control samples. Recoveries are sometimes significantly below 100 percent and depend on the lab performing the measurement and on the spiking concentration. Recovery was highest for MRI and lowest for the Casper lab and recovery decreased as the lead level in the control samples increased. Because the dust lead concentrations from field samples spanned a range that included 300 ppm found in the mid-level spike and 2100 ppm found in the high-level spike, the average measured recovery for the mid-level and high-level spike was used to adjust the dust lead measurements to correct for recovery.

All dust lead values presented in this report are corrected for recovery.

The dust measurements used in the statistical analysis and summarized in this chapter are derived from measurements made in the field and the laboratory, the area vacuumed to obtain the dust sample, the amount of lead in the dust sample, and the weight of the dust sample. The weight of the dust sample, called the "tap weight," was determined by weighing the dust that could be tapped out of the filter. Because some dust will adhere to the filter, the tap weight underestimates the true weight of the dust collected by the vacuum. Because there is no good estimate of the proportion of the dust that adheres to the filter, it was not possible to adjust the tap weight data to obtain an unbiased estimate of the dust collected by the vacuum. The laboratory procedures determined the amount of lead in both the dust that could be tapped from the filter and the dust remaining on the filter. Therefore, after applying the recovery correction discussed above to the measured amount of lead, an unbiased estimate of the amount of lead in the dust sample was obtained. The following three quantities were used in the statistical analysis:

$$\text{Dust Loading} = \frac{\text{Tap weight}}{\text{area vacummed}} \text{ in } \frac{\text{mg}}{\text{sq ft}};$$

$$\text{Lead Loading} = \frac{\text{Lead content}}{\text{area vacummed}} \text{ in } \frac{\mu\text{g}}{\text{sq ft}}; \text{ and}$$

$$\text{Lead concentration} = \frac{\text{Lead content}}{\text{Tap weight}} \text{ in ppm.}$$

Of these three measures, the dust loading is biased low and the lead concentration is biased high, because the tap weight underestimates the true dust weight in the sample. After correction for recovery, the lead loading is an unbiased estimate of the lead loading in the vacuumed area. Qualifications are added in the discussion of the results where the bias in the estimates may result in a misinterpretation.

Table 4-1 Percent recovery for dust control samples

Lab	Sample Size	Averaged measured concentration ppm	Spiked concentration ppm	Recovery (percent)	Comments
Aurora	37	36.0 ± 3.1			Unspiked samples for recovery calculation. Includes 3 high outliers
	40	347 ± 19	316.2	98.4% ± 6.1%	
	35	2027 ± 170	2097	94.9% ± 8.1%	Includes 2 low outliers
Casper	34	31.2 ± 1.1			Unspiked samples for recovery calculation
	32	314 ± 10	316.2	89.4% ± 3.2%	
	33	1754 ± 99	2097	82.2% ± 4.7%	
MRI	3	33.0 ± 12.9			Unspiked samples for recovery calculation
	3	392 ± 232	316.2	114% ± 74%	
	3	2184 ± 1890	2097	103% ± 90%	

Based on GFAA analysis of dust (soil dust) control samples. source: Analysis of Soil and Dust Samples for Lead (Pb), MRI, May 8, 1991.

Note: ± values are 95% confidence intervals for the respective parameters.

4.3 Description Of The Data

This section presents descriptive statistics for the dust lead concentrations. All dust lead concentrations in this report are mass concentrations, in ppm. First outliers in the dust data are discussed, followed by descriptive statistics.

4.3.1 Outliers

The procedures used to identify outliers in the dust lead measurements assumed the log-transformed measurements had a normal distribution and used the extreme studentized deviate test at the nominal one percent level (Section 3.2.1 includes a discussion of outliers). Outliers were identified using three different models:

(1) Assume the dust lead concentration, dust lead loading, and dust tap weight measurements from each sampling location (such as floors, window sills, window wells in the wet and dry room and floors in the entry way) have a log normal distribution. Fit a mean to the log-transformed measurements and assume that the residuals have a normal distribution.

(2) Assume the log-transformed dust lead concentration measurements have a multivariate normal distribution. Calculate the Mahalanobis distance (discussed below and in Section 3.2.1) for each dwelling unit. Assume that the cube root of the Mahalanobis distance has a normal distribution.

(3) Assume the log-transformed dust lead concentrations and dust loading measurements have a multivariate normal distribution. Calculate the Mahalanobis distance for each dwelling unit. Assume that the cube root of the Mahalanobis distance has a normal distribution.

Because there were different numbers of observations from the different types of sampling locations, the following sequential procedure were used to identify and eliminate dust lead concentration outliers:

(1) Fit the first model to the data from each type of sampling location. Remove any observations that were identified as outliers at the one percent level.

(2) Fit the second (multivariate) model using data from two sampling locations. If a pair of observations from any dwelling unit was identified as an outlier, both observations were removed from the data to be analyzed. This test was performed on all pairs of sampling locations.

(3) Fit the second model using combinations of three sampling locations. If an outlier was identified, all three observations were removed from the analysis data set.

(4) This procedure was repeated applying the multivariate model to the combinations of four, five, six, and seven sampling locations that had the most observations.

(5) Finally, the third model was applied to the dust concentration and dust loading data. Observations identified as outliers in this step or in previous steps were removed from the dust concentrations, dust tap weight, and dust loading data.

The concept behind the sequence of steps above was to proceed from simple to more complex outlier identification procedures and models, moving from one variable models to many variable models. When using the multivariate model, if a combination of measurements is identified as an outlier, it is not possible to determine which of the several measurements might be in error, therefore all measurements associated with the outlier are removed. This sequential procedure for removing outliers was adopted to minimize the number of data values removed from the data set as outliers.

As mentioned in section 3.2.1, a scatter plot of two variables typically forms a cluster of points. With three variables, the scatter plot would be a three-dimensional cloud of points in space. Mahalanobis distance measures the distance of individual observations from the center of the cluster of data points. The distance measure is a function of the standard deviations and correlations among the variables. Figure 4-1 shows a scatter plots to illustrate Mahalanobis distances.

These procedures identified 11 observations as outliers. One additional concentration measurement was removed as an outlier because it was somewhat higher than other comparable measurements and had an implausibly high lead concentration (46 percent lead by weight). The identity of these outliers are shown in Table 4.2. These outliers were removed from the data used for subsequent analyses and summary statistics because they were noticeably extreme and could have an inordinate effect on the results of the regression analysis.

4.3.2 Distribution Of The Dust Lead Concentrations

The dust lead concentrations have a skewed distribution, with many low lead concentration measurements and few relatively high measurements. The distribution can be roughly described by a log normal distribution. For the log normal distribution, the log-transformed data (that is the natural log of the measured lead concentrations) has a normal or bell shaped distribution. Figure 4-2 shows histograms of the lead concentration measurements for the entry way, drip line, and remote dust samples, respectively, using a log scale. The vertical scales are identical for all histograms, therefore the area under the histogram indicates the relative number of measurements. The histograms suggest that the window well data are slightly skewed to the right even after using the log transformation.

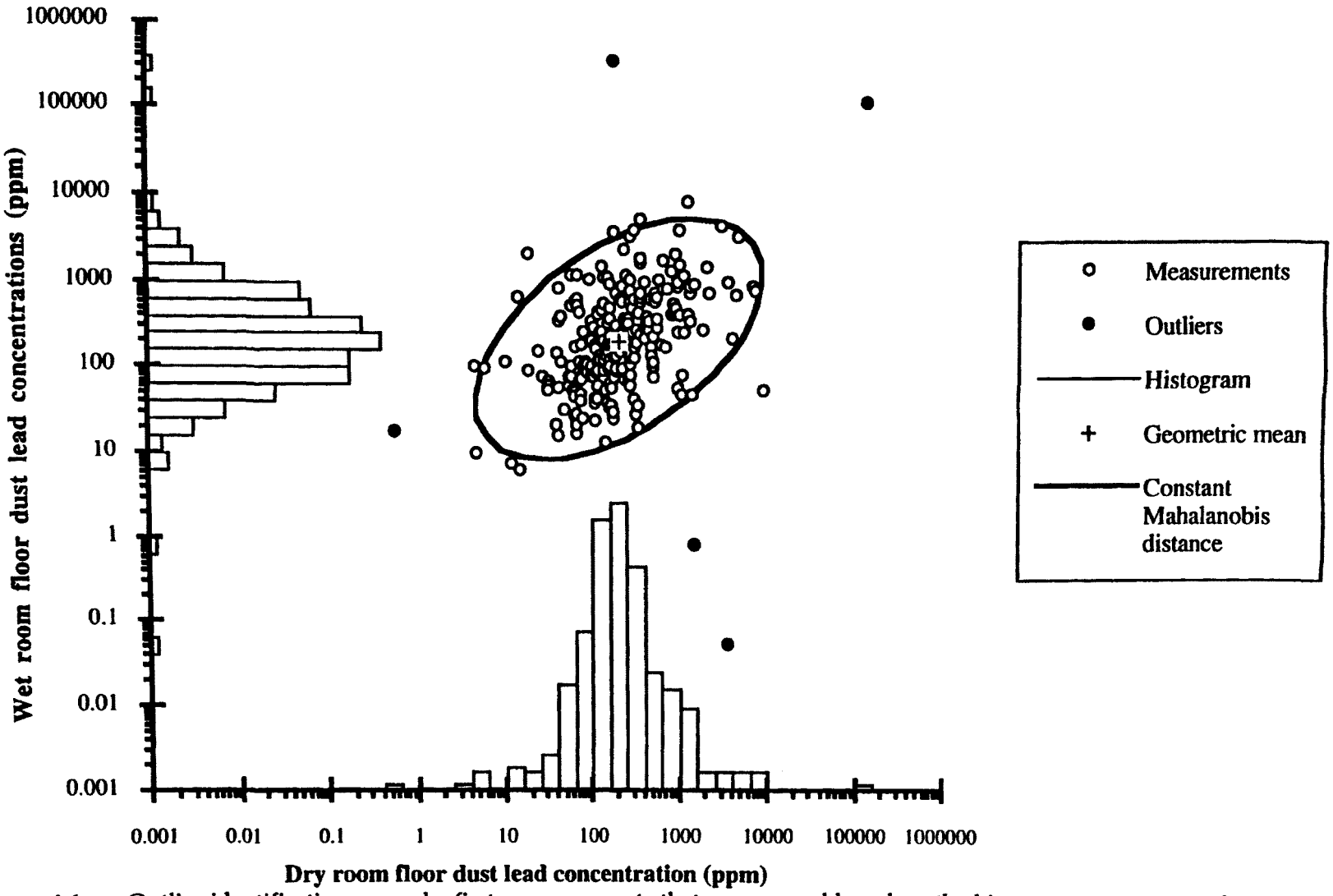


Figure 4-1 Outlier identification example: first, measurements that are unusual based on the histograms are removed, second, measurements which are unusual based on the Mahalanobis distance are removed

Table 4-2 Outliers that were removed from the dust data before analysis

Location of dust samples	Number of measurements before outlier removal	ID of the dwelling unit	Dust lead concentration (ppm)	Studentized residual [b]	p-value
Dry room floor	272	2751402	132,900	4.73	0.0006
		0430306	0.58	-4.47	0.0021
Dry room window sill	209	0520700	5,846,000	4.62	0.0008
Dry room window sill	78	1551704	457,200[a]	3.00	0.19
Entry way floor	273	0340505	1.01	-4.54	0.0015
		1731603	1.32	-4.65	0.0009
		2651206	1.94	-4.45	0.0023
		0430306	2.21	-4.47	0.0021
Wet room floor	269	1041607	0.05	-5.41	<0.0001
		0131102	354,800	5.14	0.0001
		2751402	114,300	4.61	0.0011
		2622603	0.84	-4.15	0.0088

[a] This outlier was removed even though the p-value was not significant because the dust lead concentration was implausible.
[b] The studentized residual is the number of standard deviations by which the outlier measurement differs from the mean.

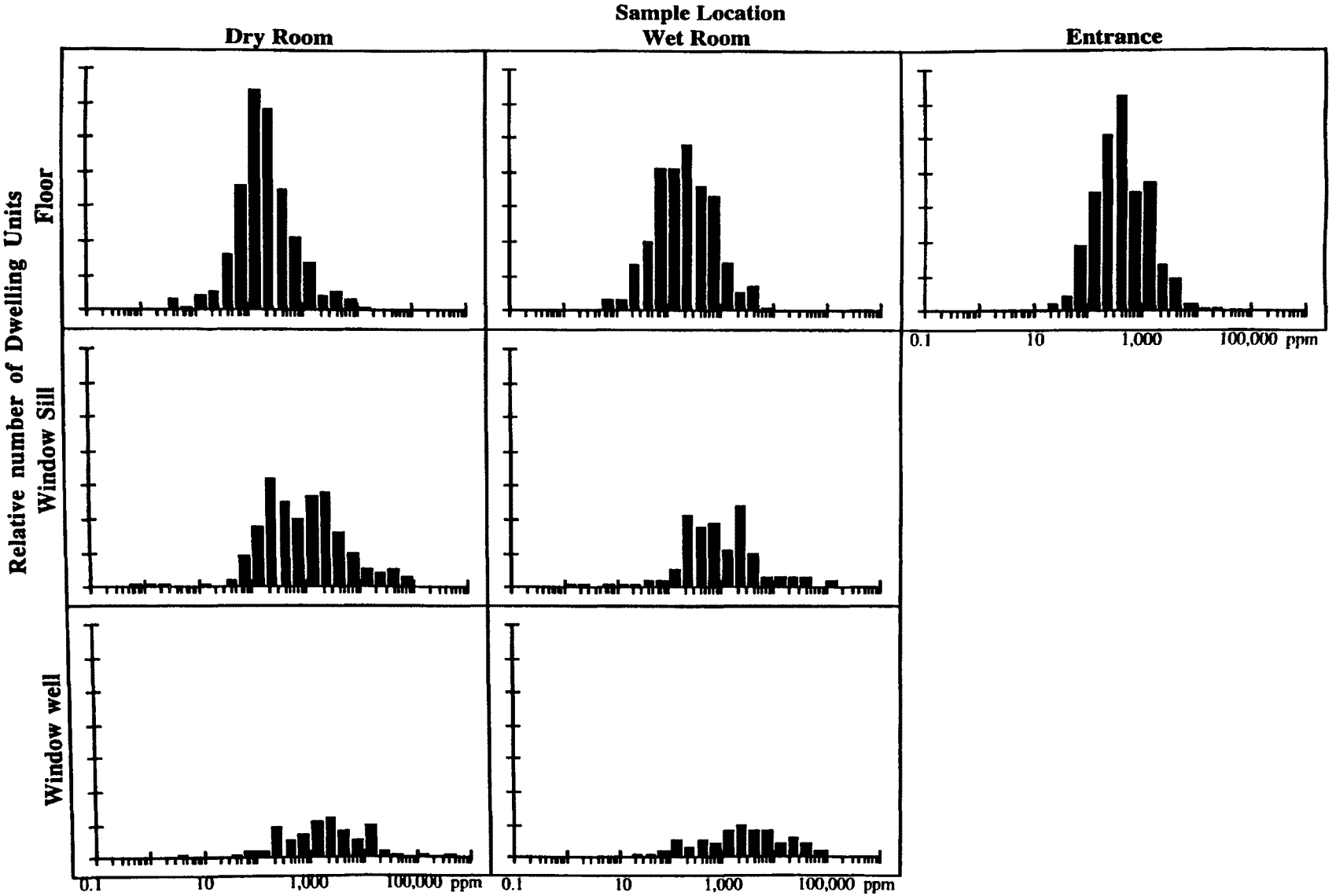


Figure 4.2 Histograms of dust lead concentrations by sampled room and sample location within the sampled room

4.3.3 Means Standard Deviations, And Descriptive Statistics

Table 4-3 presents descriptive statistics for the dust lead concentration measurements: the sample means, standard deviations, coefficients of variation, selected percentiles, geometric means, and standard deviations of the log-transformed measurements. Tables 4-4 and 4-5 show descriptive statistics for dust lead loadings and total dust loadings (loading is the amount of lead or dust on a given surface area). It is important to note that loading values strongly reflect dust accumulation since the last time a home was vacuumed. Lead concentrations (ppm) in dust, however, do not drastically change after vacuuming. For the purpose of this report, concentration data is assumed to be more informative in determining associations between different sampling locations and for determining relationships with soil lead data.

A review of the tables and figures shows that the floor dust lead concentrations are lower, on average, and less variable (as measured by the interquartile range) than the window sill and well levels. The entryway floor dust lead concentrations is higher than the other two floor dust lead levels. The two window sill samples are comparable to each other, as are the two window well samples. The highest dust lead concentrations were typically on the window wells. The fact that dust lead concentrations in window wells are significantly higher than soil and dust lead concentrations from other locations suggests that other sources, such as paint, contribute more lead to window well dust than does soil or dust from floor locations.

4.4 Interrelationships In The Dust Data

4.4.1 Differences Among Sample Locations

Within most dwelling units, floor dust samples were collected at three different locations (dry room, wet room, and entrance), and window sill and window well dust samples were collected at two different locations (dry room and wet room). For these three different types of samples (floor, window well, and window sill) the paired differences between the log-transformed measurements were used to determine if the geometric means at different locations were significantly different. For comparing measurements at different locations, only dwelling units with measurements at both locations were used. Therefore, the differences being tested may be slightly different from those shown in Tables 4-3, 4-4, and 4-5, which summarize the data for all dwelling units with the indicated type of sample.

Table 4-3 Descriptive statistics for the dust lead concentrations measurements (unweighted)

Set of data	Dry room floor	Entry way floor	Wet room floor	Dry room window sill	Dry room window well	Wet room window sill	Wet room window well
Number of measurements summarized	270	269	265	208	77	131	71
Number of outliers removed	2	4	4	1	1	none	none
Arithmetric mean (ppm) (95% confidence interval)	514 (371 to 658)	871 (671 to 1,071)	471 (366 to 575)	4,350 (2,666 to 6,034)	5,891 (2815 to 8968)	4,729 (2,236 to 7,223)	9,080 (5,398 to 12,762)
Standard deviation (ppm)	1,198	1669	864	12,318	13,554	14,426	15,555
Coefficient of variation	2.33	1.92	1.84	2.83	2.30	3.05	1.71
Percentiles (ppm)							
maximum	11,287	18,563	8,376	96,491	109,165	104,368	83,634
upper quartile	378	922	483	2,511	5,745	2,583	7,450
median	188	380	198	753	1,920	826	2,432
lower quartile	102	201	83	259	515	289	575
minimum	3	21	6	1	5	1	22
Geometric mean (ppm) (95% confidence interval)	206 (177 to 240)	428 (374 to 490)	204 (175 to 239)	853 (667 to 1,055)	1,669 (1,123 to 2,480)	906 (660 to 1,245)	2,279 (1,444 to 3,595)
Mean of the log transformed measurements	5.33	6.06	5.32	6.74	7.42	6.81	7.73
Standard deviation of the log transformed measurements	1.26	1.13	1.28	1.80	1.74	1.84	1.93

Table 4-4 Descriptive statistics for the dust lead loading measurements (unweighted)

Set of data	Dry room floor	Entry way floor	Wet room floor	Dry room window sill	Dry room window well	Wet room window sill	Wet room window well
Number of measurements summarized	273	274	275	233	84	158	74
Number of outliers removed	2	4	4	1	1	none	none
Arithmetric mean (ug/ft2) (95% confidence interval)	6.60 (4.03 to 9.17)	13.3 (8.41 to 18.3)	4.66 (2.57 to 6.74)	65.1 (28.7 to 101)	981 (9.37 to 1,952)	126 (0 to 284)	669 (362 to 976)
Standard deviation (ug/ft2)	21.5	41.5	17.6	282	4778	1001	1325
Coefficient of variation	3.26	3.11	3.77	4.33	4.56	7.95	1.98
Percentiles (ug/ft2)							
maximum	205	380	233	2,638	40,455	11,899	7,139
upper quartile	3	8	3	25	475	12	528
median	1	3	1	5	86	2	90
lower quartile	0	1	0	1	15	0	19
minimum	0	0	0	0	0	0	0
Geometric mean (ug/ft2) (95% confidence interval)	1.11 (0.90 to 1.37)	2.61 (2.13 to 3.21)	0.76 (0.61 to 0.94)	4.73 (3.45 to 6.48)	74.6 (41.5 to 134)	2.48 (1.68 to 3.65)	78.9 (42.6 to 146)
Mean of the log transformed measurements	0.11	0.96	-0.26	1.55	4.31	0.91	4.37
Standard deviation of the log transformed measurements	1.79	1.73	1.85	2.44	2.70	2.46	2.67

Table 4-5 Descriptive statistics for the dust loading measurements (unweighted)

Set of data	Dry room floor	Entry way floor	Wet room floor	Dry room window sill	Dry room window well	Wet room window sill	Wet room window well
Number of measurements summarized	269	268	264	208	77	131	71
Number of outliers removed	2	4	4	1	1	none	none
Arithmetric mean (mg/ft2) (95% confidence interval)	20.3	23.4	17.4	67.9	175	69.6	168
Standard deviation (mg/ft2)	55.0	44.3	63.5	338	301	328	319
Coefficient of variation	2.71	1.89	3.65	4.97	1.72	4.72	1.90
Percentiles (mg/ft2)							
maximum	528	351	955	3,967	2,062	3,040	2,370
upper quartile	16	24	14	42	202	16	201
median	5	7	5	8	83	4	59
lower quartile	2	2	1	2	32	1	16
minimum	0	0	0	0	0	0	0
Geometric mean (mg/ft2) (95% confidence interval)	5.54 (4.53 to 6.77)	6.58 (5.28 to 8.20)	4.13 (3.34 to 5.11)	7.54 (5.56 to 10.2)	69.9 (49.6 to 98.5)	4.23 (2.87 to 6.24)	43.0 (25.5 to 72.3)
Mean of the log transformed measurements	1.71	1.88	1.42	2.02	4.25	1.44	3.76
Standard deviation of the log transformed measurements	8.58	8.74	8.66	9.13	8.42	9.16	9.11

Figure 4-3 summarizes the location-to-location differences by type of measurement -- lead concentrations, dust loadings, and dust lead loadings. The figure shows the geometric mean ratios (with 95 percent confidence intervals) between corresponding locations in the dry and wet rooms, and between the entrance floor samples and the wet room floors. For example, it shows the dry room floor measurement relative to the measurement on the wet room floor. In Figure 4-3, a ratio confidence interval that does not include 1.0 indicates a significant difference between the wet room and the other respective location. Two confidence intervals for floor measurements that do not overlap each other also indicate significant differences. Thus, the geometric mean floor dust lead loading at the dwelling unit entrance was significantly greater than in the dry room, which was, in turn, significantly greater than in the wet room. The geometric mean floor dust lead concentration at the dwelling unit entrance was significantly greater than the wet or dry room floors. The geometric mean window sill dust lead loading in the dry room was significantly greater than in the wet room. The geometric mean floor dust loadings at the entrance and in the dry room were significantly greater than in the wet room. The geometric mean window sill dust loading in the dry room was significantly greater than in the wet room. These differences are statistically significant at the 0.001 level. No differences between the wet and dry room were significant for the window well measurements.

In less statistical terms these results can be summarized as follows:

- Except in window wells, less dust (that which could be taped out of the filter) per square foot was found in the sampled areas in the wet room than in the dry room or entrance.

- The lead represented a greater proportion of the dust in the entrance samples than in the wet room and dry room samples.

- The lead per square foot in the floor dust was greatest for the dust at the entrance, less for the dust in the dry room, and least for the dust in the wet room.

4.4.2 Correlations Between Locations

Tables 4-6 and 4-7 present the correlations among the seven dust sampling locations for both dust lead loadings and dust lead concentrations. All but two of the correlations are significant at the one percent level; most are significant at 0.1 percent level.

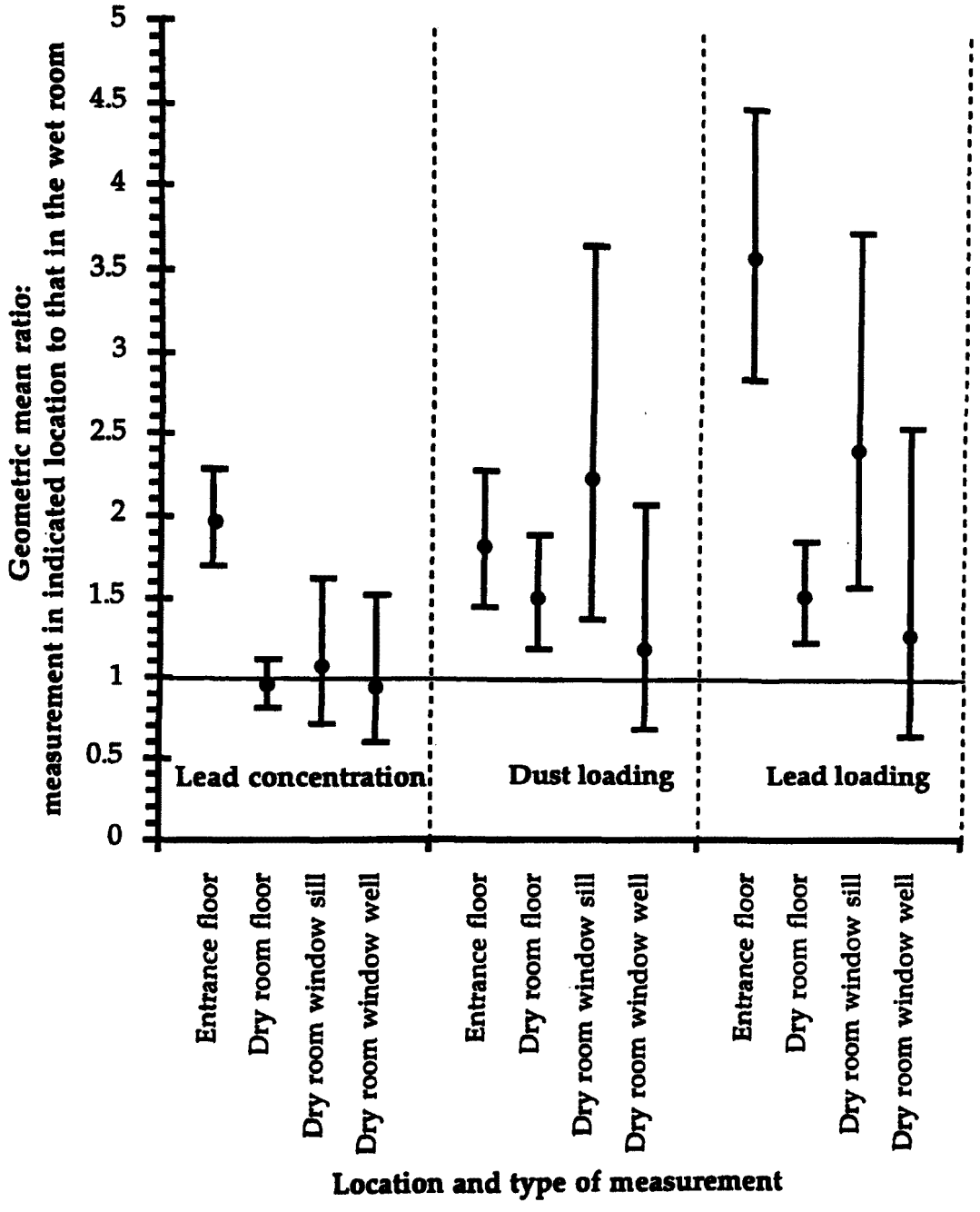


Figure 4-3 Geometric mean ratios of the dust measurements in the indicated room to corresponding locations in the wet room of the same dwelling unit with 95% confidence intervals, by type of measurement

Table 4-6 Correlations between log-transformed dust lead loading from different locations around the same dwelling unit

	Dust Lead Loading (µg/sq ft)						
	Dry room floor	Dry room window sill	Dry room window well	Entrance	Wet room floor	Wet room window sill	Wet room window well
Dry room floor		0.4846	0.3557	0.5634	0.5486	0.3891	0.1694
		0.0001	0.0011	0.0001	0.0001	0.0001	0.155
µg/sq ft	273	225	81	267	268	154	72
Dry room window sill	0.4846	1.0000	0.4774	0.4106	0.3485	0.4343	0.4412
	0.0001		0.0001	0.0001	0.0001	0.0001	0.0003
µg/sq ft	225	233	73	228	228	146	63
Dry room window well	0.3557	0.4774	1.0000	0.3501	0.3258	0.4736	0.5654
	0.0011	0.0001		0.0016	0.0026	0.0003	0.0001
µg/sq ft	81	73	83	79	83	54	45
Entry way	0.5634	0.4106	0.3501	1.0000	0.4520	0.3239	0.3855
µg/sq ft	0.0001	0.0001	0.0016		0.0001	0.0001	0.0008
	267	228	79	274	269	155	72
Wet room floor	0.5486	0.3485	0.3258	0.4520	1.0000	0.4152	0.3359
	0.0001	0.0001	0.0026	0.0001		0.0001	0.0037
µg/sq ft	268	228	83	269	275	156	73
Wet room window sill	0.3891	0.4343	0.4736	0.3239	0.4152	1.0000	0.4717
	0.0001	0.0001	0.0003	0.0001	0.0001		0.0002
µg/sq ft	154	146	54	155	156	158	59
Wet room window well	0.1694	0.4412	0.5654	0.3855	0.3359	0.4717	
	0.1550	0.0003	0.0001	0.0008	0.0037	0.0002	
µg/sq ft	72	63	45	72	73	59	74

Note: In each set of table entries, the top number is the correlation coefficient, the middle is the probability that a sample correlation this far from zero might occur by chance if there were actually no correlation in the underlying population, and the bottom number is the number of paired measurements used to calculate the correlation.

Table 4-7 Correlations between log-transformed dust lead concentrations from different locations around the same dwelling unit

	Dust Lead Concentrations (ppm)						
	Dry room floor	Dry room window sill	Dry room window well	Entrance	Wet room floor	Wet room window sill	Wet room window well
Dry room floor (ppm)		0.3481 0.0001 199	0.2530 0.0296 74	0.4459 0.0001 259	0.4626 0.0001 255	0.1582 0.0744 128	0.4905 0.0001 69
Dry room window sill (ppm)	0.3481 0.0001 199		0.5883 0.0001 59	0.3554 0.0001 199	0.3312 0.0001 196	0.2848 0.0028 108	0.5166 0.0001 56
Dry room window well (ppm)	0.2530 0.0296 74	0.5883 0.0001 59		0.5196 0.0001 73	0.3488 0.0023 74	0.3919 0.0085 44	0.6838 0.0001 40
Entry way (ppm)	0.4459 0.0001 259	0.3554 0.0001 199	0.5196 0.0001 73		0.4773 0.0001 254	0.2815 0.0012 129	0.4140 0.0004 68
Wet room floor (ppm)	0.4626 0.0001 255	0.3312 0.0001 196	0.3488 0.0023 74	0.4773 0.0001 254		0.2295 0.0100 125	0.3019 0.0130 67
Wet room window sill (ppm)	0.1582 0.0744 128	0.2848 0.0028 108	0.3919 0.0085 44	0.2815 0.0012 129	0.2295 0.0100 125		0.6986 0.0001 45
Wet room window well (ppm)	0.4905 0.0001 69	0.5166 0.0001 56	0.6838 0.0001 40	0.4140 0.0004 68	0.3019 0.0130 67	0.6986 0.0001 45	

Note: In each set of table entries, the top number is the correlation coefficient, the middle is the probability that a sample correlation this far from zero might occur by chance if there were actually no correlation in the underlying population, and the bottom number is the number of paired measurements used to calculate the correlation.

4.4.3 Measurement Variation And Differences Between Locations

This section reports on an analysis of the dust measurement variation. Conceptually, it is similar to the soil measurement variation analysis discussed in Section 3.3.3. For this discussion it is assumed that the variable of interest is the average dust lead concentration across the room's floor, window sills, or window wells.

Using lead concentration in floor samples as an example, because the floor dust sample typically was obtained from a vacuumed area of 4 square feet, its lead concentration only approximates the average over the floor. Because the dust lead concentration is likely to vary across the floor area, the actual lead concentration in the sample may be different from the average lead concentration across the floor. This difference contributes to the measurement variation. Because of uncontrolled variation in the sample preparation and final measurement step, the laboratory procedures will also contribute to the measurement variation.

As with the soil data, it is possible to estimate the measurement variance by analyzing the ratio of measurements from different locations within the dwelling unit and by making some reasonable assumptions.

We can use the ratio of the measurements at two different locations if we assume that (1) the measurement variance is the same for similar samples from different rooms, and (2) that the ratio of the measurements from different rooms is constant, independent of other factors such as the presence of lead-based paint. A preliminary analysis suggested that these assumptions are reasonable. If the assumption that other factors do not affect the difference between dust measurements is incorrect, the variance estimate will tend to be larger than the actual value.

Table 4-8 presents the estimated variance of the log-transformed measurements due to both laboratory variation and variation associated with selection of the sample in the field. These variances are for measurement of the average value across the floor, window sill, or window well. Because the estimates for the window sill and window well were similar and were not very precise due to the relatively small number of samples, the estimates for window wells and window sills were pooled.

As can be seen from Table 4-8, measurements on window sills and wells have greater variance than similar measurements on floors, and measurements of lead concentration have smaller measurement variance than measurements of lead loading. Assuming the measurement errors have a log normal distribution, 95 percent of the lead concentration measurements from floor samples are expected to be within a factor of 4.8 of the average lead concentration across the floor. For example, if a lead concentration measurement is 500 ppm, the true lead concentration across the floor is most likely (i.e. 19 out of 20 times) between 104 and 2400 ppm (500/4.8 and 500*4.8 respectively). For the measure with the greatest measurement variance, lead loadings in window sills, one measurement from this survey is likely to be within a factor of 450,000 of the average lead loading across all windows in the room. Clearly the lead loading in the vicinity of the window well and sill varies greatly from window to window. Some of this variation may be associated with the difficulty in effectively vacuuming the crevasses in some window wells.

Table 4-8 Variance of one log-transformed measurement of dust concentration, dust loading, or lead loading around the geometric mean measure for the sampled room, by sample type.

Measurement	Sample Type	
	Floor dust	Window sill or window well dust
Dust loading	1.61	2.41
Lead loading	3.06	6.64
Lead concentration	0.80	3.98

4.4.4 Factors Related To Dust Loading

Regression analysis was used to identify factors that predict dust loading. The explanatory variables used in the regression included whether the dust sample was taken on a rug or bare surface in the wet room, dry room, and entrance; paint damage in the wet room, dry room, and entrance; the presence of a rug in the entry hall (if there was an entry hall); floor level of the wet or dry room; dwelling unit age; and the county (as a classification variable). The objective of the analysis was to quickly identify the factors deemed most important in predicting dust loading. Separate regressions were performed for the different dust sampling locations.

The regressions predicting floor dust loading explained about 45 percent of the variance in the data and indicated that differences in floor dust loading between counties were very significant. These differences are likely to be due, in part, to climatic differences. Other factors that were important in some regressions included the presence of damaged paint, the floor level of the room being sampled, and the presence of a rug. The amount of dust collected in the entrance floor increased as the age of the dwelling unit increased. A similar relationship was not found for the dust amounts on the wet and dry room floors. The presence of damaged paint was positively associated with increased amount of dust; however, only in the wet room was this relationship statistically significant.

Generally more dust was collected from rugs than when sampling on a bare floor. This difference may be related to differences in either differences in the amount of dust or differences in how the vacuum sampler collected dust on the two surfaces. If an entrance hall existed, the presence of a rug in the entry hall was correlated with lower dust amounts on the dry room floor. More dust was found on the dry room floor when the sampled dry room was above the first floor.

The presence of a hall rug and the presence of damage in the wet room were found to be marginally significant for predicting dust loading on the wet room window sill. Otherwise, the regressions for predicting dust loading on window wells or sills showed no significant factors, including county-to-county differences that were so significant for the floor data. The regressions for predicting window well and sill dust loadings had fewer observations making identification of significant factors more difficult.

The differences noted above may be due to many factors that were not recorded in the survey data, in particular, the cleaning habits of the occupants. Therefore, caution is recommended when drawing conclusions from these results.

4.5 Limitations In The Data

The dust data have some outliers but otherwise have no serious statistical problems. The outliers that do exist are very unusual relative to the bulk of the observations and thus there is little question that these observations are not typical, even if the cause of the outlier cannot be identified. These outliers are identified in Table 4-2 and were removed from all analyses presented in this report. Removal of these values from other analyses that are performed on the data is also recommended. In addition, there are a few unusual observations that were not far enough out to be removed as outliers. Overall, about one percent of the measurements were removed as outliers. This small percentage suggests

that the number of incorrect observations or observations that might be unusual enough to significantly affect the results of a statistical analysis is small. Therefore, the data can be used for analysis after removing the outliers; however, results based on few observations may be less precise than the statistics would indicate due to the affect of an unusual observation that was not identified.

The results from dust samples that were too small for laboratory analysis were not reported. For the analyzed samples, none of the lead content measurements was below the relevant detection limit. Therefore, none of the statistical analyses of dust measurements in this report are based on censored data.

Although the measurement error for floor dust lead concentrations might be considered to be large (i.e., within a factor of 4.8), the measurement error is small compared to the differences in dust lead concentrations between homes and, in spite of the measurement error, significant correlations and differences between locations can be identified. The measurement errors for the window well and window sill data are larger. Although the data can be used for analysis, the size of the measurement error and the relatively smaller number of window sill, and particularly window well, samples limit the usefulness of this portion of the data.

Interpretation of the results should reflect how the samples and measurements were taken. The selection of sample locations within the selected floor, window sill, or window well was determined by the technician and was often based on the location of the greatest dust accumulations. In addition, the problems of measuring the dust weight resulted in biased estimates of dust loading and lead concentration. Therefore, for making national estimates the following general statements can be made:

- Dust and lead loading depend on the sampling equipment used. Another sampler with a different pump or nozzle might have collected more or less dust. Other studies suggest that the lead concentration measurements are not greatly affected by the choice of sampling equipment.

- Given the dust sampling equipment used, the dust loading estimates are biased; however, the direction of the bias cannot be determined (the low tap weight reduces the estimates, the sample collection from high dust areas increases the estimates).

- Given the dust sampling equipment used, the lead loading estimates are biased and tend to overestimate the average lead loading across the floor or sampled area. If the lead loading estimates are used to estimate the loading in high dust areas (perhaps a worse case), the bias of the estimates cannot be determined.

- The lead concentration estimates are biased and tend to overestimate the true lead concentration in the dust.

Since the selection of dwelling units was based on a multistage probability sample, inference from the sample to the population of dwelling units nationally should be based on a weighted analysis reflecting the sample selection weights. For the purpose of estimating the relationship between measurements or variables from the same dwelling unit, use of the sample weights is not required. To the extent that the bias in the estimates is the same for all samples, the bias discussed above will not affect the analyses discussed in subsequent chapters.

5. Analysis Of Classification Bias Using The Soil Measurements

The *Comprehensive and Workable Plan for the Abatement of Lead-Based Paint in Privately Owned Housing: Report to Congress*, reports that an estimated 18 percent of privately-owned housing units have soil lead levels in excess of 500 ppm, and a similar 18 percent have dust lead loadings in excess of federal standards. These estimates are based on classification of housing units according to their measured soil lead and dust lead levels. The soil lead measurements provide relatively unbiased estimates of the lead concentration in the top layer of soil at the locations of the soil sample. However, when the soil lead measurements are used to classify dwelling units into those with high versus low soil lead levels in soil, the estimated proportion of dwelling units with high soil lead levels may be biased (see the *Report On The National Survey Of Lead-Based Paint In Housing* for an extensive discussion and analysis of this issue for lead-paint measurements). The following discussion uses the general definition: a dwelling unit has high soil lead levels if the **arithmetic average** lead concentration across a defined area of interest[1] is greater than a specified level (which is referred to here as a soil standard). Misclassification rates based on the geometric mean lead concentration are discussed briefly.

Assume that the area of interest is to be classified based on one soil sample collected at random within the area of interest. The one soil lead measurement may be greater or less than the average soil lead concentration across the areas of interest due to both (1) variation in the soil lead concentration across the area of interest and (2) measurement variation in the measurement process at the lab. Due to both of these sources of variation, the classification of dwelling units as having high soil lead levels is imperfect. Table 5-1 shows how dwelling units might be correctly classified or misclassified. The estimate of the proportion of dwelling units with high soil lead concentrations will be biased if the number of dwellings units incorrectly classified as having high soil lead is different from the number misclassified as have low soil lead concentrations.

The estimate of the proportion of dwelling units having high soil lead concentrations can be affected by the measurement variation and the number of samples (and measurements) used for classification. The example above addresses misclassification based on one measurement. Although similar principles apply, the probability of misclassification is different when the classification is based on (1) one measurement, (2) the maximum of three measurements (for example, at the entrance, drip line, and remote locations), or (3) the average of three measurements. The following sections address the misclassification bias based on these three classification criteria.

[1]For the National survey of Lead-based Paint in Housing, the size and location of this "area of interest" has not been explicitly defined.

Table 5-1 Effect of measurement variation on the classification of dwelling units as having or not having high soil lead concentrations

Classification of the dwelling unit based on one soil sample	True classification of the dwelling unit	
	Has **low** soil lead concentrations	Has **high** soil lead concentrations
Has **low** soil lead concentrations (Soil lead measurement is less than the soil standard)	Correct classification	Incorrect classification based on the soil lead measurement Results in **too few** dwelling units classified as having high soil lead concentrations
Has **high** soil lead concentrations (Soil lead measurement is less than the soil standard)	Incorrect classification based on the soil lead measurement Results in **too many** dwelling units classified as having high soil lead concentrations	Correct classification

5.1 Classification Based On One Measurement

The probability of misclassification based on one measurement can be estimated using the following assumptions about the distribution of the log-transformed soil lead measurements:

(1) Let Y stand for the log-transformed average soil lead concentration across the area of interest and y stand for the log-transformed soil lead measurement in a randomly located soil sample within the area of interest, and e is the log-transformed measurement error. Assume $y = Y + e$.

(2) Assume Y and e have a normal distribution such that the untransformed measurement errors (e^e) have a lognormal distribution with mean of 1.0.

(3) Assume the variance of Y and e are known or can be estimated from the data.

These assumptions are consistent with the distribution of the soil lead measurements. Note that the measurement errors are assumed to be unbiased in the untransformed units because the assumed objective is to classify based on the arithmetic mean concentration across the area of interest, not the geometric mean.

Figure 5-1 shows the theoretical distribution of the true average lead concentrations (Y) and the soil lead measurements (y) for hypothetical soil samples with the same mean and variance as the actual entrance soil samples. Due to sample-to-sample variation and lab measurement variation, the soil lead measurements tend to spread out more than the actual average lead concentrations. In the untransformed units, the two distributions have the same mean.

The vertical dotted line indicates an example soil standard of 500 mg/g. The proportion of dwelling units classified as having high soil lead concentrations is the area under the solid curve to right of the standard. Using this standard, 12 percent of the dwelling units have high soil lead concentrations and 11 percent of the dwelling units are classified as having high soil lead concentrations based on the measurements. The area under the gray curve to the right of the dotted line is less than the area under the solid curve and to the right of the dotted line. Therefore, for this example, estimates of the incidence of dwelling units with high soil lead concentrations will be biased low; i.e., will, on the average, be lower than the true incidence.

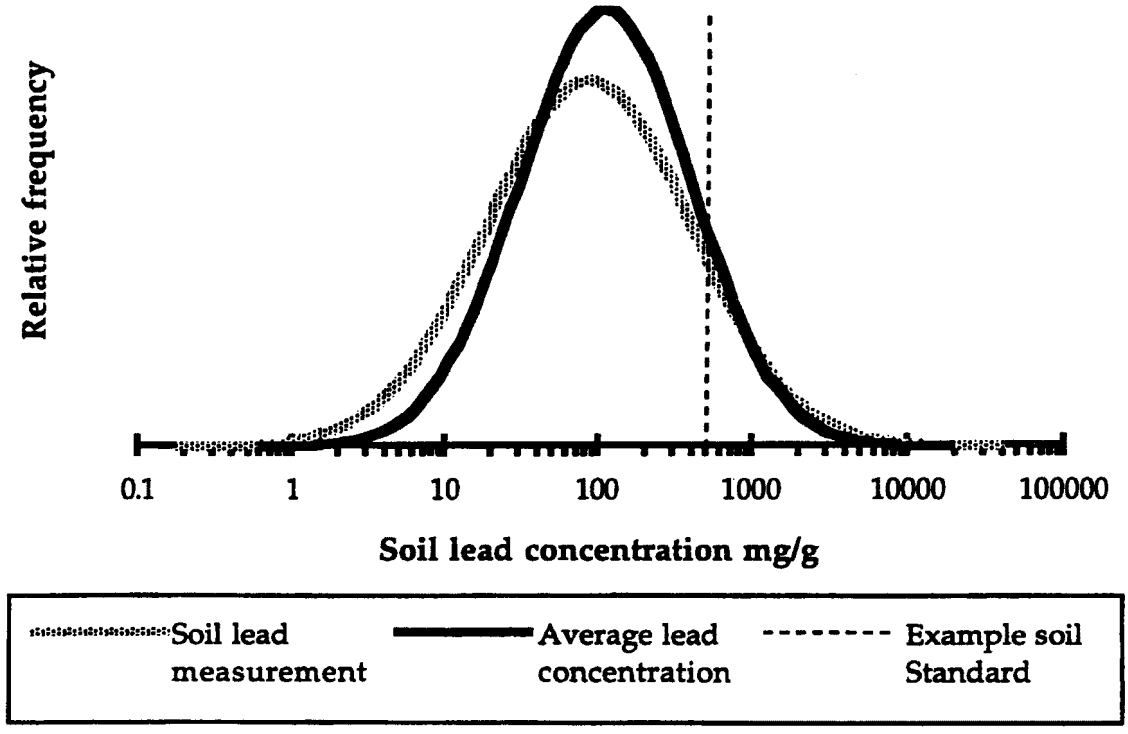


Figure 5-1 Example of misclassification due to measurement error

By calculating the percentage of dwelling units classified as having high soil lead concentrations for different soil standards, the bias in the estimates can be calculated as a function of the observed percentage. Figure 5-2 shows the expected bias in the percentage of dwelling units classified as having high soil lead concentrations based on (1) one measurement in the area of interest, (2) the maximum of three soil measurements, and (3) the average of three soil measurements. For these calculations, the soil measurements are assumed to have the same geometric mean and variance as the entrance soil measurements.

As can be seen in Figure 5-2, based on one measurement, the estimated percentage of dwelling units with high soil lead concentrations is unbiased at 0, about 8, and 100 percent. Below 8 percent, classification using one soil measurement tends to overestimate the percentage of dwelling units with high soil lead concentrations. Above 8 percent, classification using one soil measurement tends to underestimate the percentage of dwelling units with high soil lead concentrations. The estimated percentage can be off by as much as 9 percent (on the low side).

5.2 Classification Based On The Maximum Of Three Readings

When the maximum of three random samples is used to classify the soil, the probability of misclassifying a dwelling unit as having high soil lead concentrations increases. Using the assumptions presented in the last section, the bias in the percentage of dwelling units with high soil lead concentrations was calculated for classification based on the maximum of three randomly located soil samples within the area of interest. The results are shown in Figure 5-2 using the heavy dashed line.

The misclassification rates shown in Figure 5-2 are higher than those that are likely in the Lead-Based Paint survey. For Figure 5-2 we assumed that the three soil samples were randomly located within the same area of interest. In the Lead-Based Paint survey, the three soil samples were located in three different, approximately non-overlapping, areas (in the vicinity of the entrance, drip line, and remote). Because the remote samples generally showed lower concentrations than the entrance and drip line samples, the classification based on the entrance, drip line, and remote soil measurements were equivalent to classification based on fewer than three samples. This can be seen in the following example: if the remote measurements are so much lower than the entrance and drip line that the remote measurement is never the maximum of the three, then the classification will in effect be based only on the entrance and drip line measurement; i.e., the maximum of two measurements. The actual misclassification rates using the maximum of the entrance, drip line, and remote measurements is between the solid dark line and the dashed dark line.

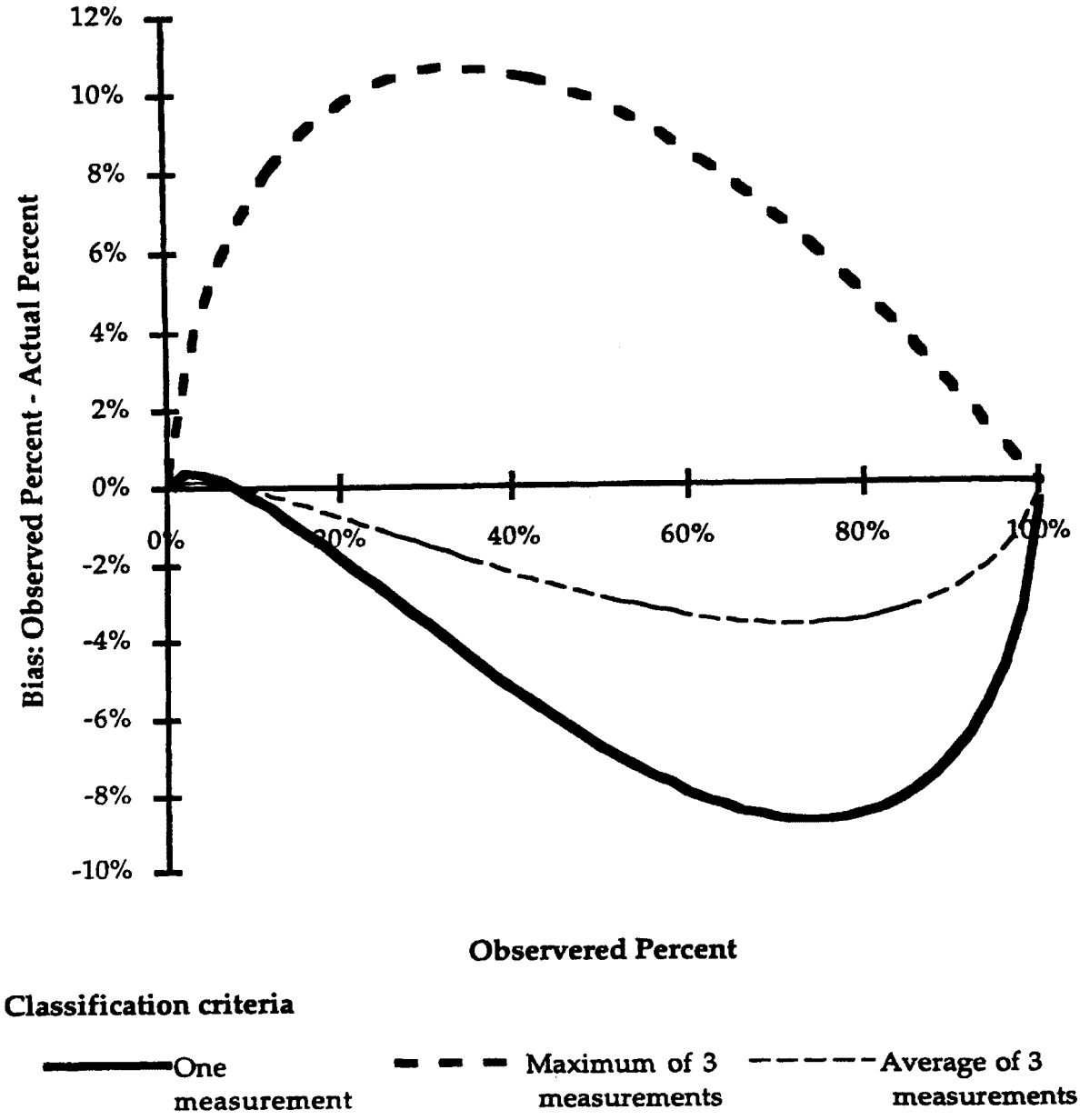


Figure 5-2 Misclassification due to soil lead measurement error

Bias in the observed percent of dwelling units classified as having high soil lead concentrations as a function of the observed percent using (1) one measurement, (2) the maximum of three measurements, or (3) the average of three measurements to classify the dwelling units.

5.3 Classification Based On The Average Of Three Readings

When the average of three random samples is used to classify the soil, the probability of misclassifying a dwelling unit as having high soil lead concentrations is closer to zero than when using one measurement. Using the assumptions presented in Section 5.1, the bias in the percentage of dwelling units with high soil lead concentrations was calculated assuming that the classification was based on the average of three randomly located soil samples. The results are shown in Figure 5-2 using the light dashed line.

5.4 Classification Using The Geometric Mean Lead Concentration

If the soil standard is defined in terms of the geometric mean lead concentration instead of the arithmetic mean, the misclassification rates will differ from those shown in Figure 5-2. Figure 5-3 shows the bias in the estimated percentage of dwelling units with geometric mean soil lead concentration greater than a standard where the classification is based on (1) one measurement, (2) the maximum of three measurements, and (3) the geometric mean of three measurements. Except when basing the classification on the maximum of three measurements, There is less classification bias when the soil standard is defined in terms of the geometric mean soil concentration rather than the arithmetic mean.

5.5 Recommendations

The discussion above presents different procedures that might be used to classify dwelling units as having high soil lead concentrations. The bias in the classification procedures depends on (1) the basis of the standard (arithmetic or geometric mean), (2) the number of samples used to classify the area of interest, (3) whether the maximum or the average is used for classification and, (4) to a lesser extent, the assumptions used to model the data. Classification based on the average (or geometric mean) of measurements from several samples across the area of interest has the least bias and is generally recommended. Classification based on the maximum of several measurements generally has positive bias (i.e., results in an overestimate of the percentage of dwelling units with high soil lead) and has the largest bias.

The geometric mean will generally be less than the arithmetic mean which will in turn be less than the maximum of several measurements. If long-term exposure is important, use of the arithmetic average is recommended. If acute exposure is important, classification based on the maximum (which in effect puts more weight on the higher measurements) would be more protective. On the other hand, the geometric mean puts more weight on the lower concentration measurements. Because high lead concentrations are generally of more concern than low concentrations, use of the geometric mean for classification is generally not recommended.

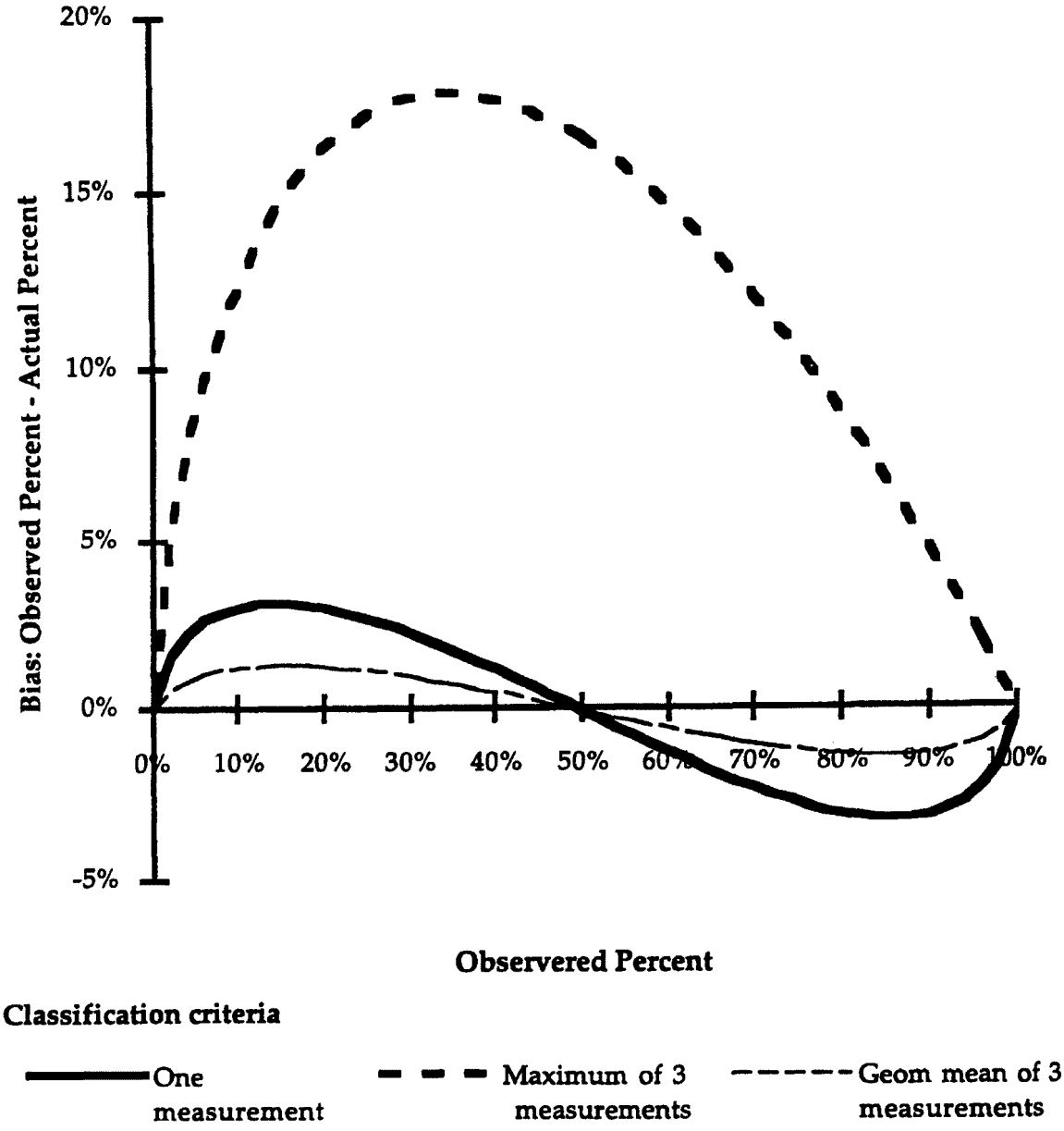


Figure 5-3 Classification bias due to soil lead measurement error when classifying using the geometric mean concentration

Bias in the observed percent of dwelling units classified as having high soil lead concentrations as a function of the observed percent using (1) one measurement, (2) the maximum of three measurements, or (3) the geometric mean of 3 measurements to classify the dwelling units

Application of the results above to the data from the National Survey depends on the definition of the "area of interest." If (1) the entire soil area around the dwelling unit is considered the area of interest, (2) the dwelling unit is to be classified based on the average soil lead concentration across the entire area, (3) the remote, drip line, and entrance samples can be considered to be either random samples across the area of interest or random samples within three strata of the same size, then use of the arithmetic average is recommended.

If three disjoint "areas of interest" are defined, viz., the areas in the vicinity of the remote, drip line, and entrance sample, and the classification is to be based on the maximum of the average soil lead concentrations across the three areas, then using the maximum of the three measurements is generally recommended. However, the bias using this classification procedure can be relatively large and will depend on the relative soil lead levels in the three areas. If it is known that all three areas have similar soil lead concentrations (relative to the magnitude of the measurement error), use of the maximum of three measurements will overestimate the actual percentage of dwelling units with high soil lead concentrations. In this case use of the average may be preferred.

Which procedure to use will depend primarily on what criteria is believed to be relevant. For example, if soil lead exposure is believed to be a major source of elevated blood lead levels and if a child is assumed to play in the soil at different locations with equal frequency, and if the arithmetic average exposure (chronic) is thought to be relevant, then using the classification based on the average soil lead concentration across the site makes sense and classification based on the average of the entrance, drip line, and remote soil samples provides a relatively unbiased classification procedure. On the other hand, if interior dust is thought to provide most of the lead exposure and interior dust appears to be most affected by the soil at the exterior entrance, classification based on only the entrance soil sample might be chosen.

6. Analysis Of Statistical Association Between Soil Lead And Lead-Based Paint And Dust

Many researchers believe that soil lead comes mainly from paint lead and automobile emissions. There is evidence in the research literature in support of this hypothesis. (The *Comprehensive and Workable Plan for the Abatement of Lead-Based Paint in Privately-Owned Housing* includes a review of the literature.) Similarly, interior dust lead is believed to come principally from paint lead and soil lead. The objective of this chapter is to explore these hypotheses through an analysis of the statistical associations among the soil, dust and paint data from the national survey.

The statistical analyses focus on regression techniques, which identify the variables that "best predict" soil lead and dust lead levels. The regression techniques also yield estimates of the predicted soil and dust lead levels as functions of the values of the predictor variables -- paint lead, for example. It is to be noted that a strong statistical association between the predictor variables and the predicted variables does not, by itself, establish a causal relationship among them. It is possible for two variables to have a strong statistical relationship, but no substantive causal relationship. Both variables may be caused by a third, unidentified variable, or the relationship may be a statistical artifact.

Section 6.1 presents the statistical model for the analyses. Section 6.2 presents the results of the regression analyses designed to predict soil lead levels. Section 6.3 presents the regressions designed to predict the dust lead hazard. The soil measurements are described in Section 3.2 and the measurement error for the soil lead measurements is discussed in detail in Section 3.3. Sections 4.3 and 4.4 discussed the distribution and measurement errors for the dust lead loadings.

6.1 Statistical Approach

In an attempt to capture the effects of local traffic volumes and changing population patterns, the national survey data was supplemented with additional data on traffic volumes in the neighborhoods of the 284 privately-owned housing units in the sample and Census population data for each of the 30 sampled counties. These data are described in the next section.

6.1.1 Traffic And census data

Because automobile emissions are thought to contribute to lead in soil, a study of traffic patterns near dwelling units sampled in the Nation Survey of Lead Based Paint in Housing was conducted to provide a measure of traffic suitable for analysis. The objective was to explore statistical correlations between lead in soil and average daily traffic volumes.

For each dwelling unit, a quarter-mile circle, centered on the dwelling unit, was drawn to scale on 1980 census tract maps. Each road intersecting the circle was noted, as was the length of the road segment inside the circle. In the event that two or more dwelling units were in close proximity to one another, those houses were included within the same circle. In some small towns (populations less than 500) in the Midwest, some houses were marked with rural delivery numbers rather than street addresses. In these cases, since it was not possible to locate dwellings on 1980 census tract maps, houses with rural delivery numbers were considered to have traffic volumes similar to those with street addresses in the same town.

An extensive effort was launched to collect data on Average Daily Traffic (ADT) counts for the streets intersecting the quarter-mile circle. An ADT is the average number of motor vehicles - automobiles, trucks, buses, etc. - that pass a given point on a road in a 24-hour period. Traffic volume maps or ADT books were usually available from state or local departments of transportation. Only current (1985-1992) ADTs were obtainable; data for earlier periods were never available. Where maps or books were not available, local government agencies provided traffic counts for roads of interest over the telephone. For roads where no counts were taken, traffic engineers advised of likely counts for each road in question. Traffic engineers frequently offered upper and lower bounds of ADT counts on roads where no counts were taken, from which a reasonably close estimate of traffic volume was drawn.

In some cases, imputations were made based on counts from nearby streets likely to have similar traffic patterns.

After ADTs were obtained, vehicle miles per day for each dwelling unit were calculated by multiplying the distance of each road within the quarter-mile circle by its average daily traffic count, and adding these products over all roads in the circle. The traffic variable, with units of vehicle miles per day, was established and used in determining the possible contributions of automobile emissions to soil lead. The distribution of the vehicle miles per day is skewed to the right, with a mean and median of 10,240 and 5,855 vehicles miles per day, respectively. The coefficient of variation is 1.25. Half of the traffic values are between 1,950 and 14,253 vehicle miles per day, with minimum and maximum values of 25 and 90,400 vehicle miles per day.

In an effort to measure past population and traffic effects, Census data from the 1920 through 1990 decennial censuses inclusive were collected. These data consist of county-wide populations and gave a rough measure of population levels when most automobiles burned leaded gasoline and lead-based paint was widely used.

6.1.2 Regressions To Identify Important Predictors Of Dust Lead And Soil Lead

Linear regression using two or more predictor variables (usually called multiple regression) can be used to both identify the variables that may be sources of lead and to estimate the relative contribution of each source of lead to the dust lead. Regression is also used to identify predictors of the soil lead concentration.

Let us assume that the true lead concentration (average lead concentration over a specified area) Y is related to three variables X_1, X_2, and X_3 as described by the following equation:

$$Y = A + B_1X_1 + B_2X_2 + B_3 X_3.$$

The Xs are called independent variables. Y is called the dependent variable because it is assumed to depend on or to be predictable from the Xs. The values A, B_1, B_2, and B_3 are parameters in the equation. Knowledge of the parameters tells us which sources of lead contribute to the potential lead hazard. In reality, not only are the parameters in this equation unknown, but so is the identification of the variables that relate to the true lead concentration.

We can use regression to fit the following equation to y, measurements of Y, and x_1, x_2, and x_3, the measurements of X_1, X_2, and X_3:

$$y = a + b_1x_1 + b_2x_2 + b_3 x_3 + e$$

The values a, b_1, b_2, b_3 , etc. are called the parameters estimates or regression coefficients. The regression procedure calculates parameter estimates that minimize the variance of the prediction errors (ε). The results of the regression analysis also include estimates of the precision of the parameter estimates.

If (1) the dependent variable, y, is the only variable that has measurement error, (2) the measurement errors, ε, are independent and the expected magnitude of the measurement error is constant, and (3) the equation used in the regression has the same independent variables and mathematical form as the true relationship between Y and the Xs, then the regression parameters a, b_1, b_2, b_3 will be unbiased estimates of the true parameters A, B_1, B_2, B_3.

Two factors can contribute to biased regression estimates and therefore to making incorrect conclusions about the sources of potential lead hazard: (1) lack of knowledge about the true relationship between the independent and dependent variables and (2) measurement error in both the independent and dependent variables. Procedures to minimize the bias in the estimates include: (1) keep all variables in the analysis that might affect the dependent variable and (2) use special techniques to estimate the parameters when the independent variables have measurement error. If important variables are left out, the estimates of y based on the remaining variables may be biased. If extra variables are included in the model, the parameter estimates for the true variables will be unbiased; however, they may not be as precise as if the extraneous parameters were not in the model.

6.1.3 A Model For The Data

The model (i.e., the equation fit using regression) was primarily derived by considering different physical processes by which the lead might move from painted surfaces, automobile exhaust, and other possible sources into soil or dust. Some additional terms were added to the model because preliminary analysis indicated that they

might be important. The following paragraphs discuss the different possible sources of lead and a theoretical basis for selecting independent variables in the model.

Consider the following possible explanation of how lead in paint may move to the floor dust and affect the dust lead loading. Abrasion may slowly move the lead from the painted surfaces to the floor. If the abrasion removes a fixed thickness of paint per unit time, the amount of lead removed with the paint will be proportional to the volume lead concentration, which is the ratio of the surface lead concentration (XRF measurement, mg/sq cm) and the thickness of the paint. The amount of abraded paint will also be proportional to the area of paint. The abrasion rate might be proportional to the number of occupants in the dwelling unit.

Most of the lead in the abraded paint will fall to the floor. The contribution to the lead loading in the dust will depend on the floor area (with larger floor area the same lead will be spread over a wider area resulting in less lead per square foot) and the time the dust has collected since the last cleaning. Both the total dust loading and the lead loading will be affected in the same way by the floor area and the time over which the dust collects. Therefore, when modeling the dust lead concentration, the time over which the dust collects and the floor area of the sampled room should not affect the dust lead concentration.

Based on the discussion above, a model for dust lead concentration due to paint abrasion would be:

$$\begin{bmatrix}\text{Dust lead} \\ \text{concentration}\end{bmatrix} = \begin{bmatrix}\text{Transport} \\ \text{rate}\end{bmatrix} \frac{\begin{bmatrix}\text{Paint lead} \\ \text{loading}\end{bmatrix}}{\text{paint thickness}} \begin{bmatrix}\text{Painted} \\ \text{surface} \\ \text{area}\end{bmatrix} \text{Occupants} \qquad \text{(eq. 1)}$$

The transport rate is the relative rate of deposition of leaded and nonleaded dust onto the floor.

A similar equation would describe the contribution of lead from other rooms in the dwelling unit with the exception that a term for the number of rooms must be added. If the paint damage results in removing chips of paint (i.e., the entire layer of paint) rather than abrasion of the surface, then the equation would not have the paint thickness term.

Taking the log transformation of both sides of equation (1) gives:

$$\begin{Bmatrix}\text{Dust lead} \\ \text{concentration}\end{Bmatrix} = \begin{Bmatrix}\text{Transport} \\ \text{rate}\end{Bmatrix} + \begin{Bmatrix}\text{Paint lead} \\ \text{loading}\end{Bmatrix} - \begin{Bmatrix}\text{paint} \\ \text{thickness}\end{Bmatrix} + \begin{Bmatrix}\text{Painted} \\ \text{surface area}\end{Bmatrix} + \{\text{Occupants}\}$$

where the curly brackets indicate that the log transformation was used.

If it is assumed that the paint lead contribution to the dust is greater when the paint is damaged, one can add another term for the increase in the paint contribution for an increase in the percent of the paint which is damaged. However, damage only increases

the lead from paint if the paint actually has lead. To model this, a term for the interaction of paint lead and damage was also added to the model.

The contribution of soil lead to the dust lead will be proportional to the number of occupants who might track soil into the dwelling unit and the lead concentration in the soil. Other factors which may affect the contribution of soil to dust include floor level on which the dust was collected and the presence of an entry hall or rug.

The final model includes terms for paint lead contributions from both the wet and dry rooms in the dwelling unit. Although the contribution of the lead from different rooms and different sources will be additive in the untransformed scale , for simplicity the model assumes they are additive on the log-transformed scale. This assumption is expected to have relatively little effect on the results, and is expected to be less important than the selection of the variables to be added to the model. However, to compensate for the possible nonlinear relationship between the log-transformed paint lead and dust lead, a quadratic term for paint lead is included in the model. To compensate for possible nonlinear relationships due to the modeling of the log-transformed data, preliminary regressions were used to identify other quadratic and interaction terms that might affect the prediction of dust and soil lead concentrations. Quadratic and interaction terms that were significant in the preliminary regressions were included in all the final regressions.

Since we do not have data on the thickness of the paint in the dwelling unit, another correlated variable was used in the model. The thickness of the paint will be proportional to the age of the dwelling unit. Therefore, the dwelling unit age was included in the model.

Factors that may affect the soil lead concentration include the traffic history in the vicinity of the dwelling unit that might have contributed lead from gasoline; the paint on and in the dwelling unit; the general condition of the dwelling unit; the same factors which affect the contribution of the paint to the dust; and background soil lead levels. The survey data do not include information on the history of traffic and background lead levels; however population levels, population growth rates, the location (county), and traffic levels in recent times may be correlated with these variables and are also included in the model.

As part of the survey, the presence of each of the following six items was noted: (1) roof, gutter, or down spout holes or damage; (2) chimney with cracks, loose bricks or chimney out of plumb; (3) unstable walls or walls with large cracks or missing boards; (4) two or more broken windows or doors; (5) porch or steps with broken or missing elements; and (6) major visible cracks or damage in the foundation. For the regressions, the general condition of the dwelling unit was measured by the number of factors noted out of this list.

Table 6-1 lists the independent variables in the final soil and dust models.

Table 6-1 Terms included in the final regression model (all are log-transformed)

Variable	Number of parameters
Paint lead loading, average XRF measurement weighted by painted surface area (mg/sq cm)	3 (wet room, dry room, exterior, and linear combinations)
Fraction damaged paint Area of damaged paint as a fraction of the area of all +paint, weighted by painted surface area, plus 1.0	3 (wet room, dry room, exterior, and linear combinations)
Area of painted surfaces (square feet)	3 (wet room, dry room, exterior, and linear combinations)
Number of wet rooms	1
Number of dry rooms	1
Soil lead concentration (ppm) (Dust regressions only)	3 (entrance, drip line, remote, and linear combinations)
Local traffic volume (miles per day)	1
Coded dwelling unit age (7 = Before 1920, 6 = 1920 to 1939, 5 = 1940 to 1949, 4 = 1950 to 1959, 3 = 1960 to 1969, and 2 = 1970 to 1979)	1
Number of occupants	1
(Number of wet rooms)^2	1
(Number of dry rooms)^2	1
(Local traffic volume)^2	1
(Paint lead concentration)^2	3 (wet room, dry room, exterior, and linear combinations)
Paint lead concentration * fraction damaged paint interaction	3 (wet room, dry room, exterior, and linear combinations)
Local traffic volume * Coded dwelling unit age interaction	1
Exterior building condition on a scale from 0 to 6 (Soil regressions only)	5
Sample surface (rug, floor, unknown) (Dust regressions only)	3
Floor level of sample location (first, above first, unknown) (Dust regressions only)	3
County code as a classification variable (Soil regressions only, used in place of the population and county area variables)	30 counties
County population (every 10 years from 1920 through 1990, used in place of the county code)	8
County area (square miles) (Used in place of the county code)	1

The model for soil lead concentrations does not include dust concentrations as independent variables because the primary sources of lead in the soil are lead paint and lead from gasoline. The dust may be on the pathway as lead moves from the paint to the soil; however, dust is not an ultimate source of lead for the soil. Lead in dust collected by the vacuum sampler generally collects over a relatively short period of time (the time between cleanings) compared to the time over which lead has accumulated in the soil. The model for dust lead concentrations includes the soil as an independent variable because, over the short period of time that the dust collects, the soil is a source of lead, rather than the traffic, which is the long-term source of lead for the soil. The traffic is included in the dust model because the interaction term of traffic by age was significant in several regressions and because greater traffic may contribute to or be correlated with greater disturbance of soil dust.

6.1.4 Selection Of The Independent Dust Variable

The regression analysis can be used to model either the dust lead loading or the dust lead concentration. Blood lead levels in young children are generally believed to be more correlated to lead loading than to lead concentration. On this basis, predictions of the lead loading would be more useful for assessing the lead hazard to young children. For the purposes of identifying sources of lead in dust, either regression models for the lead loading or the lead concentration can be used.

Preliminary regressions indicated that the best predictor of dust lead loading is the weight of the dust sample. In other words, more lead collects on the floor as more dust collects on the floor. These results indicate that one important way to reduce dust lead loading is to remove the dust.

The lead concentration in the dust will depend on the relative rates of deposition of leaded dust and nonleaded dust and should be relatively independent of the amount of dust collected. However, the measured lead concentration may not be independent of the amount of dust if either 1) the periodic floor cleaning is more (or less) efficient at picking up nonleaded dust than leaded dust, or 2) different proportions of the leaded and nonleaded dust are removed from the sample filter by tapping.

Preliminary regressions were used to test whether the measured dust lead concentration and dust loading were independent. If the log-transformed dust loading and dust lead concentration are independent and if each has the same measurement error, their difference and sum will also be independent. Note that the sum of the log-transformed dust loading and lead concentration is the log of the lead loading. In a regression predicting the sum from the difference (and other terms), the coefficient for the difference will be zero if the dust loading and lead concentration are independent. The results of these preliminary regressions are consistent with the assumption that the measured lead concentration is independent of the dust loading.

Assuming the dust loading and lead concentration are independent and ignoring measurement error, the results of regressions on either the lead loading or the lead concentration should provide the same information about the factors, other than the time since the dust was last cleaned up, that affect the lead content in the dust.

Measurement errors in the independent variables affect the parameter estimates. For theoretical reasons discussed above, predictions of floor lead loading should include two more independent variables (the floor area in the sampled rooms, and the dust loading) than for predicting the lead concentration. Increasing the number of independent variables with error tends to increase the bias in the parameter estimates associated with the measurement error. This would argue in favor of modeling the lead concentration.

Given that the regressions could model either the lead loading or the lead concentration, an important argument for modeling the lead concentration is that some of the floor area measurements are missing in the survey records. Including the floor area into the regression equation reduces the number of dwelling units that can be used in the regression.

Therefore, the results presented in Sections 6.2 and 6.3 are for regressions predicting the dust lead concentration. Section 6.4 compares the regression results using both lead concentration and lead loading as independent variables to illustrate the effect of changing the model on the parameter estimates.

6.1.5 Use Of Linear Combinations Of The Independent Variables

High correlations between independent variables reduce the precision of the corresponding regression estimates. Linear combinations of the independent variables that are uncorrelated can, in some cases, provide insight into the factors that affect the dependent variables. For example, the paint lead (XRF) measurements for the dry room, wet room, and exterior surfaces are highly correlated with each other. As a result, it is difficult to discern which of these sources contribute lead to the dust. Three alternate variables, (1) the average of the log-transformed paint lead measurements, (2) the difference between the wet and dry room log-transformed paint lead measurements, and (3) the difference between the interior and exterior log-transformed paint lead measurements are less correlated with each other and can be used to determine if the average lead paint measurements might be related to the dust measurements. These three alternate variables were identified as having low correlations based on principle components analysis. Although these alternate variables are only one set of many that might be considered, they have the advantage that they are relatively easy to interpret.

The regressions were performed using three combinations of the variables collected in the National Survey:

1) Variables based on how the samples were collected in the survey (e.g., the log-transformed area weighted average paint lead concentration (XRF) in the wet room, dry room, and exterior surfaces, or the soil lead concentrations at the entrance, drip line, and remote locations).

2) Linear combinations based on principle components that have low correlations among themselves (e.g., the average of the log-transformed XRF for the wet room, dry room, and exterior surfaces, the difference between the XRF in the wet and dry room and between the interior (i.e., average of the log-transformed XRF from the wet and dry rooms) and the exterior). For the soil concentrations, the principle components corresponded to the average of the log-transformed

concentrations, difference between the entrance and drip line, and the difference between the remote and close (average of the log-transformed drip line and entrance lead concentrations).

3) Linear combinations that had terms for the interior and exterior and difference between the wet and dry room values for the dust measurements and terms for the close, remote, and difference between the entrance and drip line for the soil lead concentrations. These linear combinations are referred to as the Interior versus Exterior in Tables 6-8 to 6-17.

The interpretation of the regression results were based on the results of all three regressions.

6.1.6 Interpretation Of The Regression Estimates

The regression estimates measure the ratio of the change in the dependent variable associated with a change in the independent variable. For example, suppose the parameter estimate for paint lead in the dry room when predicting dry room floor dust is 0.23. An increase in the dependent variable (paint lead in the dry room) by 0.23 percent is associated with an increase in the dry room floor dust concentration by 1 percent.

If (1) the independent and dependent variables measure lead concentration (for example in paint or soil or traffic, which is assumed to be proportional to the lead contribution from gasoline), (2) the model includes all factors that affect the dependent variable, (3) it can be assumed that the lead in the dependent variable is caused by the independent variables, and (4) the model correctly reflects the measurement error in all the variables, then the regression estimates can be interpreted as the proportion of lead in the dependent variable that comes from the independent variable. For example, if the regression estimate for traffic is 0.18 and the estimate for paint lead is 0.07 when predicting entrance dust, it may be reasonable to conclude that 18 percent of the lead in the dust comes from traffic related sources and 7 percent from paint lead, and the remaining 75 percent from other sources.

Since these assumptions cannot be shown to be correct, the regression estimates can provide only some indication of the relative importance of different sources of lead to the dependent variable.

6.1.7 Error In Variable Regression

Measurement error in the independent variables can affect both the regression estimates and the precision of those estimates. Although regression can determine the parameters that best predict the observed data, application of these parameters to predict lead concentrations for other data sets of homes will result in biased predictions.

Unfortunately, the procedures for obtaining parameter estimates when the independent variables are subject to error are both more difficult to use than standard multiple regression, and depend on estimates of measurement error in each variable which are usually not available. These estimates of measurement error include measurement error associated with (1) measurement instrument error, (2) random selection of the sample for measurement, and (3) error in the specification of the measurement process. An example of the third source of measurement error is the following: If the true determinant of entrance dust lead concentration is the soil lead concentration in the top half centimeter of soil, but the soil sampling procedures selected the top two centimeters of soil, then the difference in the soil concentrations in the top two centimeters versus the top half centimeter contributes to the measurement error.

Due to the relative complexity of the estimation procedure, only two simple regression models (applied to the floor dust lead concentration and the remote soil lead concentration) are presented below to illustrate the differences in the parameters that can be obtained when the error assumptions are changed.

The following steps can be used to estimate the regression parameters when there is error in all the variables:

(1) Determine the measurement error in all the variables.

(2) Divide each variable by its measurement error to scale the variable.

(3) Calculate the principle components for the covariance matrix for all of the variables (both the independent and the dependent variables).

(4) Obtain the coefficients for the least significant principle component.

(5) Divide the coefficient for each independent variable by the negative of the coefficient for the dependent variable to obtain the coefficient for the scaled variable.

(6) Multiply the coefficient for the scaled variable by the corresponding measurement error from step (1) to determine the coefficient for each independent variable.

For the soil lead concentrations where the model includes a classification effect (county), the steps above were applied to the residuals from a oneway analysis of variance on each of the variables by county.

Equations for the standard errors of these parameters are both complicated and approximate and were not calculated for these examples.

To illustrate the effect of different error assumptions on the parameter estimates, the following simplified model was fit to the floor dust lead concentration using both linear regression and principle components:

$$\begin{Bmatrix}\text{Floor dust lead}\\ \text{concentration}\end{Bmatrix} = \begin{Bmatrix}\text{Interior XRF}\\ \text{Measurement}\end{Bmatrix} + \begin{Bmatrix}\text{Interior paint}\\ \text{percent damage}\end{Bmatrix} + \begin{Bmatrix}\text{Close-in soil lead}\\ \text{concentration}\end{Bmatrix}$$

Table 6-2 shows the assumed measurement error and the parameter estimates for each variable when predicting the dry room floor dust lead, calculated from both linear regression and the principle components of the covariance matrix. The results for two linear regression models are shown, the simple model shown above, the full regression model discussed in other sections of this chapter. The differences in the parameter estimates are due to the effect of different sets of independent variables and different error assumptions.

The following simplified model was fit to the remote soil lead concentration using both linear regression and principle components:

$$\begin{Bmatrix}\text{Remote soil lead}\\ \text{concentration}\end{Bmatrix} = \begin{Bmatrix}\text{Average XRF}\\ \text{Measurement}\end{Bmatrix} + \begin{Bmatrix}\text{Local traffic}\\ \text{volume}\end{Bmatrix} + \begin{Bmatrix}\text{Dwelling}\\ \text{unit age}\end{Bmatrix}$$

Table 6-3 shows the assumed measurement error and the parameter estimates for each variable when predicting the dry room floor dust lead, calculated from both linear regression and the principle components of the covariance matrix. The table shows the results for two linear regression models, the simple model shown above, and the full regression model discussed in other sections of this chapter.

The error assumptions can have significant affects on some of the parameters. Therefore, caution is advised when making conclusions about the relative contribution of different lead sources to the dust and soil lead content.

6.1.8 Correlations

Correlations between variables can be used to identify pairs of variables for which the linear relationship is statistically significant. If two variables are significantly correlated, high (or low) values of one are associated with high (or low) values of the other. If one variable causes or directly affects another in a linear manner, the two variables will be correlated. Although causation usually implies correlation, correlation does not imply a causal relationship. Even if there is a causal relationship, correlations cannot be used to determine which variable is the cause and which the effect. In many cases, significant correlations are associated with a third, perhaps unmeasured, variable that affects the two correlated variables.

Table 6-2 Parameter estimates for dry room floor dust lead concentrations using models with different error assumptions

Variable	Measurement error in log transformed scale	Parameter estimates with 95% confidence intervals		
		Multiple regression		Principle components
		Simple model for lead concentration	Full model for lead concentration [a]	Simple model for lead concentration
Dry room floor dust lead concentration (dependent)	0.8			
Paint lead concentration averaged across wet rooms and dry rooms	0.35	0.10 ± 0.08	0.12 ± 0.10	0.09
Percent damage averaged across wet rooms and dry rooms	0.01	6.92 ± 5.63	8.50 ± 13.9	8.92
Number of wet and dry rooms	0.2	0.89 ± 0.50	0.48 ± 0.58 (wet rooms) 0.49 ± 0.54 (dry rooms)	2.90
Geometric average drip line and entrance soil lead concentration	0.5	0.25 ± 0.13	0.25 ± 0.22	0.38

[a]Parameters estimates for other variables are not shown.

Table 6-3 Parameter estimates for remote soil lead concentrations using models with different error assumptions

Variable	Measurement error in log transformed scale	Parameter estimates with 95% confidence intervals		
		Multiple regression		Principle components
		Simple model for lead concentration	Full model for lead concentration (a)	Simple model for lead concentration
Remote soil lead concentration (dependent)	0.8			
Average of the paint lead concentration in wet rooms, dry rooms, and exterior surfaces	0.35	0.12 ± 0.07	0.16 ± 0.09	0.10
Local traffic volume	0.01	0.29 ± 0.13	0.24 ± 0.14	0.47
Dwelling unit age	0.2	1.28 ± 0.41	1.10 ± 0.49	1.35

[a]Parameters estimates for other variables are not shown.

For example, if soil and dust lead concentrations are correlated, this may be due to either (1) dust lead comes from soil, (2) soil lead comes from dust, (3) the lead in soil and dust comes from a third source such as exterior paint or automobile exhaust, or (4) some combination of these factors.

Correlations measure the strength of the linear relationship between variables. If the relationship is not linear, the correlation may not be significant even though there is a causal relationship between the variables. Often a transformation of a variable can be used to make a nonlinear relationship more linear. Transformations can also be used to change the relative measurement variability of the values being correlated. For the correlations shown in this chapter, a log transformation has been used for all variables.

Table 6-4 displays the correlations for the independent variables used in the regressions to predict the dust lead concentrations. Table 6-5 displays the correlations for the independent variables used in the regressions to predict the soil lead concentrations. Table 6-6 displays the correlations of the independent variables in the regressions with themselves. In all three of these correlation tables the top number in each cell is the correlation coefficient, the middle number is the probability that a sample correlation this far from zero might occur by chance if there were actually no correlation in the underlying population, and the bottom number is the number of paired measurements used to calculate the correlation.

Table 6-4 Correlations of dust lead concentrations with independent variables used in the regression equations.

	Dust lead concentration (ppm)						
	Dry room floor	Dry room window sill	Dry room window well	Entry way	Wet room floor	Wet room window sill	Wet room window well
Paint lead	0.1869	0.2305	0.1878	0.1422	0.1277	0.1844	0.1557
dry room	0.0020	0.0008	0.1020	0.0196	0.0377	0.0350	0.1948
mg/sq cm	270	208	77	269	265	131	71
Paint lead	0.1601	0.2933	0.3530	0.1599	0.1109	0.3046	0.3485
wet room	0.0084	0.0001	0.0016	0.0086	0.0715	0.0004	0.0029
mg/sq cm	270	208	77	269	265	131	71
Paint lead	0.0762	0.0313	0.0811	-0.0132	0.0615	0.1784	0.0899
exterior	0.2122	0.6533	0.4831	0.8295	0.3184	0.0415	0.4558
mg/sq cm	270	208	77	269	265	131	71
Paint damage	0.0876	0.0384	-0.0313	0.0497	0.1047	-0.0026	0.0504
dry room	0.1560	0.5857	0.7910	0.4219	0.0926	0.9771	0.6807
percent	264	204	74	263	259	129	69
Paint damage	0.1667	0.0722	0.0907	0.1011	0.2008	0.0664	-0.0027
wet room	0.0069	0.3063	0.4391	0.1034	0.0012	0.4581	0.9827
percent	262	203	75	261	256	127	68
Paint damage	0.1349	0.0667	0.1227	0.0589	0.1030	0.1504	0.0878
exterior	0.0327	0.3541	0.3045	0.3526	0.1057	0.0969	0.4867
percent	251	195	72	251	248	123	65
Paint area	0.0822	-0.0375	-0.0275	0.0492	0.1367	0.1971	0.0473
dry room	0.1941	0.6032	0.8187	0.4378	0.0314	0.0289	0.7081
sq ft	251	195	72	251	248	123	65
Paint area	0.0608	0.0354	-0.3380	-0.0209	0.0869	-0.0719	-0.1600
wet room	0.3259	0.6155	0.0032	0.7364	0.1632	0.4182	0.1925
sq ft	263	204	74	262	259	129	68
Paint area	0.0598	-0.0694	-0.1646	-0.0356	0.0453	-0.1955	-0.0774
exterior	0.3347	0.3251	0.1583	0.5671	0.4702	0.0276	0.5306
sq ft	262	203	75	261	256	127	68
Number of	0.0342	-0.1570	-0.0296	-0.0379	-0.0793	0.0490	-0.2831
dry rooms	0.5972	0.0323	0.8120	0.5586	0.2249	0.5982	0.0258
	241	186	67	240	236	118	62
Number of	0.0191	-0.0129	0.0264	-0.0439	0.1925	-0.0607	-0.0122
wet rooms	0.7723	0.8647	0.8344	0.5086	0.0037	0.5267	0.9243
	232	176	65	229	226	111	63
Dripline soil	0.3095	0.3636	0.3784	0.4087	0.3478	0.2913	0.3924
lead conc.	0.0001	0.0001	0.0015	0.0001	0.0001	0.0015	0.0016
ppm	238	182	68	237	238	116	62
Entrance soil	0.2854	0.3331	0.4608	0.4519	0.3414	0.3468	0.3263
lead conc.	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0080
ppm	248	192	73	248	247	121	65
Remote soil	0.2135	0.2603	0.4856	0.3067	0.2956	0.3243	0.4337
lead conc.	0.0008	0.0003	0.0001	0.0001	0.0001	0.0003	0.0003
ppm	243	186	71	242	239	118	64
Dwelling	0.2914	0.2799	0.5569	0.3392	0.2812	0.2273	0.4345
unit age	0.0001	0.0001	0.0001	0.0001	0.0001	0.0090	0.0002
	270	208	77	269	265	131	71
Traffic	0.0763	0.1024	0.1822	0.1417	0.0797	0.2284	0.3551
Vehicle miles	0.2117	0.1410	0.1128	0.0201	0.1961	0.0087	0.0024
per day	270	208	77	269	265	131	71
Number of	0.0870	-0.0019	0.0153	0.0650	0.1337	0.1447	-0.0052
Occupants	0.1542	0.9779	0.8951	0.2884	0.0296	0.0991	0.9654
	270	208	77	269	265	131	71

Table 6-5 Correlations of soil lead concentrations with independent variables used in the regression equations

	Soil lead measurement (ppm)		
	Entrance	Dripline	Remote
Paint lead	0.2312	0.2432	0.1773
dry room	0.0002	0.0001	0.0047
mg/sq cm	260	249	253
Paint lead	0.2679	0.3103	0.3087
wet room	0.0001	0.0001	0.0001
mg/sq cm	260	249	253
Paint lead	0.2770	0.2740	0.2736
exterior	0.0001	0.0001	0.0001
mg/sq cm	260	249	253
Paint damage	0.0024	-0.0052	0.0150
dry room	0.9701	0.9361	0.8149
percent	254	243	247
Paint damage	0.1605	0.1130	0.1393
wet room	0.0107	0.0800	0.0292
percent	252	241	245
Paint damage	0.1466	0.1333	0.1873
exterior	0.0209	0.0399	0.0033
percent	248	238	244
Painted area	0.0351	0.0630	0.0529
dry room	0.5780	0.3280	0.4083
sq ft	254	243	247
Painted area	0.0820	0.0913	0.1136
wet room	0.1944	0.1578	0.0758
sq ft	252	241	245
Painted area	0.1344	0.1575	0.1509
exterior	0.0344	0.0150	0.0184
sq ft	248	238	244
Number of	-0.1387	-0.1043	-0.0552
dry rooms	0.0274	0.1057	0.3883
	253	242	246
Number of	-0.2186	-0.1816	-0.1259
wet rooms	0.0005	0.0047	0.0491
	252	241	245
Dwelling	0.5841	0.5900	0.5339
unit age	0.0001	0.0001	0.0001
	260	249	253
Traffic	0.2026	0.2375	0.2805
Veh. mi/day	0.0010	0.0002	0.0001
	260	249	253
Number of	0.0658	0.1115	0.0553
Occupants	0.2906	0.0790	0.3809
	260	249	253

Table 6-6 Correlations of the independent variables used in the regression equations with themselves

Independent Variable		Independent Variable Numbers							
		1	2	3	4	5	6	7	8
1	Paint lead dry room mg/sq cm		0.6463 0.0001 284	0.3700 0.0001 284	-0.1626 0.0066 278	0.0384 0.5261 275	0.0450 0.4653 265	0.0755 0.2097 278	0.0340 0.5742 275
2	Paint lead wet room mg/sq cm	0.6463 0.0001 284		0.3594 0.0001 284	-0.1486 0.0132 278	0.0795 0.1889 275	0.0593 0.3362 265	0.0577 0.3381 278	0.0011 0.9858 275
3	Paint lead exterior mg/sq cm	0.3700 0.0001 284	0.3594 0.0001 284		-0.0104 0.8631 278	0.1595 0.0080 275	0.1045 0.0897 265	0.0018 0.9759 278	0.0936 0.1216 275
4	Paint Damage dry room percent	-0.1626 0.0066 278	-0.1486 0.0132 278	-0.0104 0.8631 278		0.3587 0.0001 271	0.1863 0.0025 261	-0.0315 0.6009 278	0.0291 0.6336 271
5	Paint Damage wet room percent	0.0384 0.5261 275	0.0795 0.1889 275	0.1595 0.0080 275	0.3587 0.0001 271		0.1983 0.0014 258	0.0206 0.7354 271	0.0895 0.1389 275
6	Paint Damage exterior percent	0.0450 0.4653 265	0.0593 0.3362 265	0.1045 0.0897 265	0.1863 0.0025 261	0.1983 0.0014 258		0.0804 0.1956 261	0.1508 0.0153 258
7	Painted Area dry room sq ft	0.0755 0.2097 278	0.0577 0.3381 278	0.0018 0.9759 278	-0.0315 0.6009 278	0.0206 0.7354 271	0.0804 0.1956 261		0.2151 0.0004 271
8	Painted Area wet room sq ft	0.0340 0.5742 275	0.0011 0.9858 275	0.0936 0.1216 275	0.0291 0.6336 271	0.0895 0.1389 275	0.1508 0.0153 258	0.2151 0.0004 271	
9	Painted Area exterior sq ft	0.0246 0.6904 265	-0.0030 0.9616 265	0.1852 0.0025 265	0.0423 0.4960 261	0.1221 0.0500 258	0.0669 0.2777 265	0.0269 0.6657 261	0.1147 0.0658 258
10	Number of dry room	0.0374 0.5349 277	-0.0592 0.3267 277	0.1337 0.0261 277	0.0044 0.9425 277	0.0102 0.8680 270	-0.0524 0.4005 260	-0.0605 0.3155 277	-0.1137 0.0621 270
11	Number of wet room	-0.0836 0.1670 275	-0.1006 0.0958 275	0.0235 0.6975 275	-0.1231 0.0429 271	-0.1115 0.0648 275	-0.1108 0.0758 258	0.0487 0.4244 271	-0.0283 0.6402 275
12	Dripline soil lead conc. ppm	0.2432 0.0001 249	0.3103 0.0001 249	0.2740 0.0001 249	-0.0052 0.9361 243	0.1130 0.0800 241	0.1333 0.0399 238	0.0630 0.3280 243	0.0913 0.1578 241
13	Entrance soil lead conc. ppm	0.2312 0.0002 260	0.2679 0.0001 260	0.2770 0.0001 260	0.0024 0.9701 254	0.1605 0.0107 252	0.1466 0.0209 248	0.0351 0.5780 254	0.0820 0.1944 252
14	Remote soil lead conc. ppm	0.1773 0.0047 253	0.3087 0.0001 253	0.2736 0.0001 253	0.0150 0.8149 247	0.1393 0.0292 245	0.1873 0.0033 244	0.0529 0.4083 247	0.1136 0.0758 245
15	Dwelling unit age	0.1424 0.0164 284	0.1739 0.0033 284	0.1482 0.0124 284	0.1437 0.0165 278	0.2005 0.0008 275	0.2154 0.0004 265	0.0138 0.8186 278	0.1500 0.0128 275
16	Traffic veh miles/day	-0.0894 0.1331 284	0.0589 0.3226 284	0.0056 0.9252 284	-0.0824 0.1705 278	-0.0516 0.3940 275	-0.0392 0.5248 265	0.0756 0.2087 278	-0.0323 0.5937 275
17	Number of Occupants	0.0589 0.3225 284	0.11774 0.0474 284	0.11211 0.0592 284	-0.01466 0.8077 278	0.01934 0.7495 275	0.07321 0.2349 265	0.04438 0.4611 278	-0.13164 0.0291 275

Table 6-6 Correlations of the independent variables used in the regression equations with themselves (continued)

Independent Variable	Independent variable Numbers continued								
	9	10	11	12	13	14	15	16	17
1 Paint lead	0.0246	0.0374	-0.0836	0.2432	0.2312	0.1773	0.1424	-0.0894	0.0589
dry room	0.6904	0.5349	0.1670	0.0001	0.0002	0.0047	0.0164	0.1331	0.3225
mg/sq cm	265	277	275	249	260	253	284	284	284
2 Paint lead	-0.0030	-0.0592	-0.1006	0.3103	0.2679	0.3087	0.1739	0.0589	0.1177
wet room	0.9616	0.3267	0.0958	0.0001	0.0001	0.0001	0.0033	0.3226	0.0474
mg/sq cm	265	277	275	249	260	253	284	284	284
3 Paint lead	0.1852	0.1337	0.0235	0.2740	0.2770	0.2736	0.1482	0.0056	0.1121
exterior	0.0025	0.0261	0.6975	0.0001	0.0001	0.0001	0.0124	0.9252	0.0592
mg/sq cm	265	277	275	249	260	253	284	284	284
4 Paint Damage	0.0423	0.0044	-0.1231	-0.0052	0.0024	0.0150	0.1437	-0.0824	-0.0147
dry room	0.4960	0.9425	0.0429	0.9361	0.9701	0.8149	0.0165	0.1705	0.8077
percent	261	277	271	243	254	247	278	278	278
5 Paint Damage	0.1221	0.0102	-0.1115	0.1130	0.1605	0.1393	0.2005	-0.0516	0.0193
wet room	0.0500	0.8680	0.0648	0.0800	0.0107	0.0292	0.0008	0.3940	0.7495
percent	258	270	275	241	252	245	275	275	275
6 Paint Damage	0.0669	-0.0524	-0.1108	0.1333	0.1466	0.1873	0.2154	-0.0392	0.0732
exterior	0.2777	0.4005	0.0758	0.0399	0.0209	0.0033	0.0004	0.5248	0.2349
percent	265	260	258	238	248	244	265	265	265
7 Painted Area	0.0269	-0.0605	0.0487	0.0630	0.0351	0.0529	0.0138	0.0756	0.0444
dry room	0.6657	0.3155	0.4244	0.3280	0.5780	0.4083	0.8186	0.2087	0.4611
sq ft	261	277	271	243	254	247	278	278	278
8 Painted Area	0.1147	-0.1137	-0.0283	0.0913	0.0820	0.1136	0.1500	-0.0323	-0.1316
wet room	0.0658	0.0621	0.6402	0.1578	0.1944	0.0758	0.0128	0.5937	0.0291
sq ft	258	270	275	241	252	245	275	275	275
9 Painted Area		0.0925	0.1427	0.1575	0.1344	0.1509	0.1619	-0.0633	0.1797
exterior		0.1369	0.0219	0.0150	0.0344	0.0184	0.0083	0.3048	0.0033
sq ft	265	260	258	238	248	244	265	265	265
10 Number of	0.0925		0.4900	-0.1043	-0.1387	-0.0552	-0.1756	-0.2007	0.3189
dry room	0.1369		0.0001	0.1057	0.0274	0.3883	0.0034	0.0008	0.0001
	260	277	270	242	253	246	277	277	277
11 Number of	0.1427	0.4900		-0.1816	-0.2186	-0.1259	-0.2017	-0.1610	0.1351
wet room	0.0219	0.0001		0.0047	0.0005	0.0491	0.0008	0.0075	0.0250
	258	270	275	241	252	245	275	275	275
12 Dripline soil	0.1575	-0.1043	-0.1816		0.7148	0.6780	0.5900	0.2375	0.1115
lead conc.	0.0150	0.1057	0.0047		0.0001	0.0001	0.0001	0.0002	0.0790
ppm	238	242	241	249	246	243	249	249	249
13 Entrance soil	0.1344	-0.1387	-0.2186	0.7148		0.6090	0.5841	0.2026	0.0658
lead conc.	0.0344	0.0274	0.0005	0.0001		0.0001	0.0001	0.0010	0.2906
ppm	248	253	252	246	260	247	260	260	260
14 Remote soil	0.1509	-0.0552	-0.1259	0.6780	0.6090		0.5339	0.2805	0.0553
lead conc.	0.0184	0.3883	0.0491	0.0001	0.0001		0.0001	0.0001	0.3809
ppm	244	246	245	243	247	253	253	253	253
15 Dwelling	0.1619	-0.1756	-0.2017	0.5900	0.5841	0.5339		0.2187	-0.0008
unit age	0.0083	0.0034	0.0008	0.0001	0.0001	0.0001		0.0002	0.9890
	265	277	275	249	260	253	284	284	284
16 Traffic	-0.0633	-0.2007	-0.1610	0.2375	0.2026	0.2805	0.2187		0.11628
veh miles/day	0.3048	0.0008	0.0075	0.0002	0.3010	0.0001	0.0002		0.0503
	265	277	275	249	260	253	284	284	284
17 Number of	0.1797	0.31886	0.13513	0.11153	0.06578	0.05532	-0.00082	0.11628	
Occupants	0.0033	0.0001	0.025	0.079	0.2906	0.3809	0.989	0.0503	
	265	277	275	249	260	253	284	284	284

6.1.9 Steps Used To Derive The Regression Equation

The following series of data processing steps were followed in order to derive the regression results and conclusions from the survey data:

(1) Identify and remove outliers from the dust and soil data. Twelve outliers were removed from the dust measurements. (See Section 4.3.1.)

(2) Based on theoretical arguments, identify terms that are likely to predict the dust lead concentration, dust lead loading, and soil lead concentration. (See section 6.1.3.)

(3) Identify additional quadratic and interaction terms that might also be related to lead concentration and loading.

(4) Perform preliminary analyses to identify (a) any remaining outliers or particularly influential observations and (b) any quadratic terms or interaction terms that are significant.

(5) Remove any outliers or influential observations that result in a significant interaction or quadratic term that would not be significant with the influential observation removed (See Table 6-7 for a list of five dwelling units removed from the regressions.)

(6) Select the independent variable for the dust regressions, either the dust lead concentration or the dust lead loading. (See Section 6.1.4.)

(7) Select the variables for the final model, including those variables thought to be predictors of the independent variable based on theoretical observations and those quadratic and interaction terms that are significant for predicting at least one of the independent variables (after removing influential observations shown in Table 6-7). These terms in the model are listed in Table 6-1.

(8) Run the regression model using three different linear combinations of the lead paint and soil variables.

(9) Run the soil model with population variables instead of the county classification variable.

(10) Summarize the results and draw conclusions.

Table 6-7 Dwelling units removed from the regressions because both (a) they were influential in making a quadratic or interaction term significant and (b) with their removal, the term was not significant.

ID of the dwelling unit	Reason for removing the dwelling unit from the regressions
1751304	This dwelling unit made the term for (number of wet rooms)^2 very significant in the dry room window sill dust regression.
1333806	This dwelling unit made the term for (dry room paint lead)^2 significant in the wet room floor dust regression.
0951004	This dwelling unit made the term for (traffic)^2 and dry room paint lead by damage interaction very significant in the drip line soil regression and (exterior paint lead)^2 very significant in the entrance soil regression.
2441509	This dwelling unit made the term for dry room paint lead by damage interaction significant in the remote soil regression.
3011509	This dwelling unit made the term for exterior painted area by dwelling unit age very significant in the remote soil regression.

Note: significant means $p < 0.05$, "very significant" means $p < 0.01$.

6.2 What Predicts Soil Lead?

This section presents the results of the analysis of the predictors of exterior soil lead concentrations. Soil lead levels were regressed on housing unit paint lead loadings, percentage of damaged paint, and areas of surfaces covered with paint; dwelling unit age and other descriptors of the housing unit; local traffic volumes, a county indicator, and 1920-1990 decennial Census populations. Tables 6-8 through 6-10 present the parameter estimates for the drip line, entrance, and remote soil regressions. For first order parameters, the table shows the parameter estimates with 95 percent confidence intervals. Terms that are statistically significant at the 5 percent level are shown in bold type. For the quadratic and interaction terms, only the statistically significant terms are listed with the parameter estimates. At the bottom of these tables are the number of observations used in the regressions and the adjusted R-square for the basic regressions, using the county code, and regressions where the county code variable is replaced by the population and county area variables.

The strongest predictors of soil lead are dwelling unit age and county of residence, for all three soil sample locations. Both are highly significant ($p < 0.01$). Soil lead concentrations increase with dwelling unit age. Dwelling unit age measures the length of time since the construction of the building and, in most cases, the last major disturbance of the soil. Thus, dwelling unit age measures the length of time lead deposits -- from whatever source -- have been accumulating on the soil. The county of residence measures the differences between counties (regional effect) and was added to the statistical model as a class variable. This county or regional effect may be due to many factors, including general population density and background soil levels of a given area.

Since paint lead was tested in three locations--wet room, dry room, and one exterior wall, each location was examined separately and in different linear combinations with each other. This procedure provided some, but limited, insight into the primary source of the paint lead measured in the soil. The parameter estimates suggest that paint lead from dwelling surfaces contribute more to the entrance and drip line soil lead samples than to the remote sample. This finding was expected because entrance and drip-line samples are closer to painted structures than are remote samples. Both interior (the average of wet room and dry room paint lead area concentrations) and exterior paint lead contribute in a statistically significant way to entrance soil lead, while the drip line soil lead is more strongly associated with the exterior paint lead and the average of the exterior, wet room, and dry room paint lead concentrations.

Local traffic volumes were also examined, and are statistically significant only for the remote samples, even though all three soil samples are highly correlated with each other. Additional factors, such as paint lead from the house, could overwhelm the traffic effects for samples collected near buildings (entrance and drip-line samples). The significant positive estimates for the square of the traffic volume for the drip line and entrance samples suggests that a contribution of lead from traffic may be significant at higher traffic volumes. The significant positive interaction between traffic volume and building age for the drip line and remote samples is consistent with increasing lead content is gasoline going back in time.

Table 6-8 Parameter estimates for drip line soil regressions

		Observed Variables	Approx. principle components	Interior versus Exterior
Paint lead (X)	Dry rooms	0.02±0.08		
	Wet rooms	0.08±0.09		
	Exterior	0.07±0.06		0.07±0.06
	Average		0.18±0.10	
	Wet - dry		0.02±0.07	0.02±0.07
	Int - Ext		-0.01±0.06	
	Interior			0.08±0.09
Proportion damaged paint (D)	Dry rooms	-4.1±9.6		
	Wet rooms	-1.7±4.5		
	Exterior	-0.3±3.2		-0.8±3.3
	Average		0.2±7.8	
	Wet - dry		-0.4±5.0	0.5±6.0
	Int - Ext		0.6±3.3	
	Interior			0.0±8.5
Painted surface area (A)	Dry rooms	-0.16±0.21		
	Wet rooms	-0.06±0.22		
	Exterior	0.03±0.14		0.03±0.14
	Average		-0.19±0.33	
	Wet - dry		0.03±0.16	0.04±0.16
	Int - Ext		-0.10±0.14	
	Interior			-0.19±0.30
Other terms	# of wet rooms	-0.57±0.56	-0.33±0.57	-0.41±0.57
	# of dry rooms	0.00±0.51	-0.06±0.53	-0.05±0.53
	Local traffic (Traf)	0.08±0.15	0.09±0.15	0.09±0.16
	Unit age (Age)	1.55±0.54	1.68±0.55	1.61±0.55
	Number of Occupants	0.34±0.32	0.35±0.32	0.38±0.32
Significant interactions and quadratic terms		Traf^2=0.15 Xext^2=0.02 Xwet*Dwet=3.0 Traf*age=0.37	Traf^2=0.15 Xavg*Davg=2.5 Traf*age=0.39	Traf^2=0.15 Xext^2=0.02 Xext*Dext=0.9 Traf*age=0.38
County differences (p-value)		0.0008	0.0029	0.0023
Exterior condition (p-value)		0.4089	0.5636	0.4943
Adjusted R-square		0.588	0.569	0.571
Number of dwelling units, n		223	223	223
Adjusted R-square using population instead of counties		0.564		

Table 6-9 Parameter estimates for entrance soil regressions

		Observed Variables	Approx. principle components	Interior versus Exterior
Paint lead (X)	Dry rooms	0.07±0.08		
	Wet rooms	0.07±0.08		
	Exterior	0.05±0.05		0.05±0.05
	Average		0.23±0.09	
	Wet - dry		0.00±0.07	0.00±0.07
	Int - Ext		0.03±0.05	
	Interior			0.12±0.09
Proportion damaged paint (D)	Dry rooms	-8.5±9.4		
	Wet rooms	1.1±4.3		
	Exterior	0.1±3.1		-0.1±3.1
	Average		-1.0±7.3	
	Wet - dry		0.7±4.7	3.8±5.8
	Int - Ext		-1.3±3.1	
	Interior			-4.1±8.2
Painted surface area (A)	Dry rooms	-0.09±0.20		
	Wet rooms	-0.07±0.20		
	Exterior	0.09±0.13		0.10±0.13
	Average		-0.05±0.30	
	Wet - dry		-0.03±0.15	-0.02±0.15
	Int - Ext		-0.12±0.13	
	Interior			-0.10±0.28
Other terms	# of wet rooms	-0.54±0.52	-0.31±0.52	-0.52±0.52
	# of dry rooms	-0.13±0.45	-0.16±0.46	-0.11±0.46
	Local traffic (Traf)	-0.10±0.15	-0.07±0.15	-0.08±0.15
	Unit age (Age)	1.20±0.51	1.35±0.50	1.24±0.51
	Number of Occupants	0.15±0.30	0.12±0.30	0.16±0.30
Significant interactions and quadratic terms		Traf^2=0.12 Xext^2=0.02	Traf^2=0.13 Xint-ext^2=0.01	Traf^2=0.12 Xext^2=0.02
County differences (p-value)		0.0029	0.0023	0.0071
Exterior condition (p-value)		0.4923	0.6959	0.6134
Adjusted R-square		0.530	0.522	0.514
Number of dwelling units, n		233	233	233
Adjusted R-square using population instead of counties		0.518		

Table 6-10 Parameter estimates for remote soil regressions

		Observed Variables	Approx. principle components	Interior versus Exterior
Paint lead (X)	Dry rooms	-0.03±0.08		
	Wet rooms	0.09±0.08		
	Exterior	0.05±0.06		0.04±0.05
	Average		0.16±0.09	
	Wet - dry		0.06±0.06	0.06±0.06
	Int - Ext		0.00±0.05	
	Interior			0.07±0.08
Proportion damaged paint (D)	Dry rooms	-4.6±8.9		
	Wet rooms	-1.6±4.3		
	Exterior	0.4±3.0		0.3±3.0
	Average		-3.9±7.1	
	Wet - dry		1.5±4.6	0.9±5.5
	Int - Ext		-2.1±3.0	
	Interior			-3.3±7.7
Painted surface area (A)	Dry rooms	-0.21±0.20		
	Wet rooms	0.02±0.20		
	Exterior	0.02±0.13		0.01±0.13
	Average		-0.19±0.30	
	Wet - dry		0.07±0.14	0.09±0.14
	Int - Ext		-0.07±0.13	
	Interior			-0.14±0.28
Other terms	# of wet rooms	-0.35±0.52	-0.25±0.51	-0.31±0.50
	# of dry rooms	0.18±0.48	0.14±0.48	0.16±0.48
	Local traffic (Traf)	0.26±0.14	0.24±0.14	0.25±0.14
	Unit age (Age)	0.95±0.51	1.10±0.49	1.00±0.50
	Number of Occupants	0.02±0.29	0.03±0.29	0.04±0.29
Significant interactions and quadratic terms		Xext^2=0.02 Traf*age=0.31	Xwet-dry^2=0.01 Xint-ext^2=0.01	Xwet-dry^2=0.01 Xext^2=0.02
County differences (p-value)		0.0032	0.0017	0.003
Exterior condition (p-value)		0.0103	0.0089	0.0084
Adjusted R-square		0.492	0.501	0.500
Number of dwelling units, n		229	229	229
Adjusted R-square using population instead of counties		0.435		

There is some indication, from the damage by paint lead interaction, that a combination of increased exterior paint damage in conjunction with higher exterior paint lead loadings is associated with increased drip line soil lead concentrations. Otherwise, the percent of damaged paint, painted surface area, and the number of rooms are almost never significant.

The exterior condition of the building was significant for predicting the lead concentration at the remote location, but not at locations close to the dwelling unit. In general, greater damage to the structure was associated with increased soil lead concentrations.

For all three soil measurements, the regressions using the county of residence as a classification variable had higher predictive power than the regressions using the population and county area, as judged by the adjusted r-square. Therefore the parameters from the regressions using population and county area are not presented. In all three regressions, the county area was statistically significant.

6.3 What Predicts Dust Lead?

This section presents the results of the analysis of the potential household lead hazard, as measured by interior dust lead levels. Interior dust lead levels -- for all seven dust sample locations -- were each regressed on housing unit paint lead concentrations, percentage of damaged paint, and surface areas covered with paint; dwelling unit age and other descriptors of the housing unit; and all three soil sample concentrations.

6.3.1 Floor Dust Lead?

A review of Tables 6-11 through 6-13 indicates that exterior soil lead contributes (statistically) to floor dust lead. It appears that the soil lead contribution to dust lead comes mainly from the two "close in" samples, that is, those at the drip line and main entrance.

Paint lead appears to contribute mainly to the dry and wet room floor dust and not to the entrance dust. However, the combination of damaged paint and high paint lead concentrations on exterior surfaces is associated with higher entrance dust lead concentrations.

There is some indication, from the damage terms or the damage by paint lead interaction, that increased paint damage is associated with increased dust lead concentrations. Although the estimates for the percent damaged exterior paint are not significant, higher entrance dust concentrations are positively and significantly predicted by a combination of higher exterior paint loadings and higher percentages of exterior paint damage. The percent of damaged paint on interior surfaces is positively related to the floor dust in the wet and dry room; however, the parameters are not all statistically significant.

Table 6-11 Parameter estimates for the dry room floor dust regressions

		Observed Variables	Approx. principle components	Interior versus Exterior
Paint lead (X)	Dry rooms	0.07±0.10		
	Wet rooms	0.04±0.10		
	Exterior	-0.02±0.06		-0.02±0.06
	Average		0.09±0.10	
	Wet - dry		-0.03±0.08	-0.02±0.08
	Int - Ext		0.05±0.06	
	Interior			0.12±0.10
Proportion damaged paint (D)	Dry rooms	4.8±16.4		
	Wet rooms	0.5±4.8		
	Exterior	1.3±3.7		1.0±3.6
	Average		9.3±9.5	
	Wet - dry		-4.0±6.0	-3.1±8.1
	Int - Ext		2.0±4.6	
	Interior			8.5±13.9
Painted surface area (A)	Dry rooms	-0.10±0.22		
	Wet rooms	0.06±0.23		
	Exterior	0.02±0.15		0.02±0.15
	Average		0.02±0.33	
	Wet - dry		0.08±0.17	0.07±0.17
	Int - Ext		-0.02±0.15	
	Interior			-0.04±0.30
Soil (S)	Entrance	0.13±0.19		
	Drip line	0.12±0.18		
	Remote	-0.09±0.20		-0.09±0.20
	Average		0.14±0.22	
	Entrace-drip line		-0.01±0.15	0.00±0.15
	Close-remote		0.13±0.18	
	Close			0.25±0.22
Other terms	# of wet rooms	0.44±0.59	0.50±0.60	0.48±0.58
	# of dry rooms	0.49±0.54	0.44±0.54	0.49±0.54
	Local traffic (Traf)	0.08±0.14	0.05±0.13	0.06±0.14
	Unit age (Age)	0.48±0.64	0.46±0.65	0.52±0.64
	Number of Occupants	-0.13±0.38	-0.09±0.38	-0.12±0.38
Significant interactions and quadratic terms		Drooms^2=0.86	Drooms^2=0.87	Drooms^2=0.86
Rug differences		0.4955	0.5129	0.5552
Floor level differences		0.3868	0.3438	0.4066
Adjusted R-square		0.161	0.160	0.161
Number of dwelling units, n		207	207	207

Table 6-12 Parameter estimates for the wet room floor dust regressions

		Observed Variables	Approx. principle components	Interior versus Exterior
Paint lead (X)	Dry rooms	0.06±0.08		
	Wet rooms	-0.04±0.09		
	Exterior	-0.02±0.06		-0.02±0.06
	Average		0.02±0.10	
	Wet - dry		-0.04±0.07	-0.04±0.07
	Int - Ext		0.03±0.05	
	Interior			0.03±0.08
Proportion damaged paint (D)	Dry rooms	-1.6±10.2		
	Wet rooms	5.9±4.7		
	Exterior	-1.2±2.8		-0.8±2.7
	Average		3.0±7.8	
	Wet - dry		1.5±5.3	3.3±6.2
	Int - Ext		1.7±3.6	
	Interior			1.2±9.2
Painted surface area (A)	Dry rooms	-0.04±0.19		
	Wet rooms	0.26±0.20		
	Exterior	0.09±0.13		0.09±0.13
	Average		0.31±0.29	
	Wet - dry		0.13±0.15	0.14±0.15
	Int - Ext		0.01±0.13	
	Interior			0.22±0.27
Soil (S)	Entrance	0.18±0.17		
	Drip line	0.08±0.16		
	Remote	0.10±0.18		0.09±0.18
	Average		0.36±0.19	
	Entrace-drip line		0.06±0.13	0.05±0.13
	Close-remote		0.03±0.16	
	Close			0.26±0.20
Other terms	# of wet rooms	0.60±0.52	0.60±0.51	0.55±0.52
	# of dry rooms	0.18±0.48	0.18±0.48	0.18±0.48
	Local traffic (Traf)	0.01±0.12	0.04±0.12	0.02±0.12
	Unit age (Age)	-0.12±0.58	-0.13±0.59	-0.10±0.58
	Number of Occupants	0.11±0.32	0.07±0.32	0.09±0.32
Significant interactions and quadratic terms				
Sampled surface (p-value)		0.0157	0.0086	0.0073
Floor level (p-value)		0.0076	0.0099	0.0101
Adjusted R-square		0.271	0.258	0.262
Number of dwelling units, n		208	208	208

Table 6-13 Parameter estimates for the entrance floor dust regressions

		Observed Variables	Approx. principle components	Interior versus Exterior
Paint lead (X)	Dry rooms	-0.05±0.08		
	Wet rooms	0.04±0.08		
	Exterior	-0.04±0.05		-0.04±0.05
	Average		-0.04±0.08	
	Wet - dry		0.03±0.07	0.03±0.07
	Int - Ext		0.03±0.05	
	Interior			-0.02±0.08
Proportion damaged paint (D)	Dry rooms	1.2±8.8		
	Wet rooms	-0.2±3.7		
	Exterior	-1.6±2.5		-1.9±2.4
	Average		3.1±6.0	
	Wet - dry		-3.7±4.3	-1.3±5.1
	Int - Ext		2.8±2.9	
	Interior			1.9±7.5
Painted surface area (A)	Dry rooms	-0.06±0.19		
	Wet rooms	0.11±0.18		
	Exterior	-0.03±0.12		-0.03±0.12
	Average		0.04±0.26	
	Wet - dry		0.08±0.14	0.08±0.14
	Int - Ext		0.03±0.11	
	Interior			0.04±0.24
Soil (S)	Entrance	0.30±0.15		
	Drip line	0.11±0.14		
	Remote	-0.07±0.16		-0.07±0.16
	Average		0.35±0.17	
	Entrace-drip line		0.10±0.12	0.10±0.12
	Close-remote		0.19±0.15	
	Close			0.41±0.17
Other terms	# of wet rooms	0.15±0.47	0.22±0.46	0.15±0.46
	# of dry rooms	0.20±0.42	0.21±0.42	0.22±0.42
	Local traffic (Traf)	-0.02±0.11	0.00±0.11	-0.02±0.11
	Unit age (Age)	0.23±0.51	0.16±0.51	0.24±0.51
	Number of Occupants	-0.04±0.29	-0.06±0.29	-0.02±0.29
Significant interactions and quadratic terms		Xext*Dext=0.8	Xavg^2=0.03 Xavg*Davg=2.3	Xint^2=0.02 Xext*Dext=0.9
Sampled surface (p-value)		0.3692	0.3209	0.3555
Adjusted R-square		0.255	0.267	0.261
Number of dwelling units, n		206	206	206

Paint surface areas, dwelling unit age, number of occupants, and the number of rooms are almost never significant.

The relative magnitude of the soil and paint parameters suggests that soil sources contribute more lead to the floor dust than do paint sources. However, the paint lead term in the model for predicting dust lead in dry rooms is larger and more significant than in models to predict dust lead concentrations in wet rooms or at dwelling entrances.

For the samples from wet room floors, the dust lead concentrations for samples collected on rugs were significantly lower than for samples collected on bare floors. In addition, the dust lead concentrations for samples collected in rooms on the second or higher floors were significantly lower than for samples collected in first floor rooms. Similar relationships were not found for samples collected in the dry room and dwelling unit entrance.

Overall, the regressions explained only 25 percent or less of the variance in the lead concentration measurements from floor samples, indicating that the regression equations have little predictive power. There are likely to be other factors that explain the variation in the dust concentrations that were not measured as part of the survey. Although the overall regression models provided a highly significant fit to the floor dust concentration data, caution is recommended when interpreting the results.

To summarize the results for the regressions on floor dust:

- The parameter estimates are consistent with the conclusion that soil contributes more lead to floor dust than does paint lead.
- Paint damage is associated with a significant increase in the interior (wet and dry room) dust lead concentrations and in the entrance dust lead concentrations in the presence of high exterior paint lead concentrations.
- The parameter estimates are consistent with the conclusion that lead in dust at the entrance to the dwelling unit comes almost exclusively from exterior soil sources.

6.3.2 Window Sill Dust Lead?

A review of Tables 6-14 and 6-15 indicates different patterns in the parameters for the wet and dry rooms. Paint lead from the wet room is a significant predictor of both dry and wet rooms window sill lead concentrations. However, this conclusion for the dry room window sills should be interpreted with caution because the signs of the parameters for paint damage and painted surface area are negative, which is opposite of what would be expected if paint lead provides a significant contribution to the window sill dust concentrations. Although the dry room data indicates a very significant link between the soil close to the dwelling unit and the dry room window sill dust concentrations, the regression results suggest that soil provides no lead to the wet room window sill lead. As with the floor samples from the wet room, window sill lead concentrations for samples from the wet room are lower when collected in rooms on the second or higher floors than from rooms on the first floor. A window sill, as discussed in Section 1.2 of this report, is the lower interior ledge of a window.

Although statistically significant, the regressions for predicting window sill dust lead concentrations explain only about one quarter of the variance in the data, leaving much of the variation unexplained. Caution in interpreting the results is recommended.

6.3.3 Window Well Dust Lead?

As shown in Tables 6-16 and 6-17, only one parameter in the window well regressions is statistically significant (higher wet room window well dust associated with lower exterior paint lead loadings) and this result is difficult to interpret because the sign is opposite of what was expected. These regressions had fewer observations than for the other locations, providing less data from which to make conclusions. Since the overall regressions were generally not statistically significant, the findings do not permit any assessment of the sources of wet or dry room well dust lead levels. A window well, also discussed in Section 1.2 of this report, is the lower part of a window between the window and the screen.

As can be seen in these tables, the parameter estimates are generally quite close and within the 95 percent confidence intervals. Therefore, the conclusions that might be obtained from regressions predicting either the dust lead loading or the dust lead concentrations are likely to be the same.

6.4 The Effect On The Soil Regression Estimates Of Using Different Independent Variables

The dust loading, rather than the dust lead concentration analyzed in Sections 6.2 and 6.3, is usually considered to be more closely related to the lead risk to small children. In order to illustrate the effect of using lead loading rather than lead concentration as the independent variable in the dust regressions, this section presents the parameters for the main effects in the dust models for the dry room floor, dry room window sill, and entrance dust using two models: (1) the model for the dust concentration discussed above and (2) a model for the dust loading with terms for the dust weight per unit area and floor area in the sampled room (if applicable) added. Tables 6-18, 6-19, and 6-20 show the regression parameters for the main effects (those excluding the quadratic and interaction terms) for the two models applied to the dry room floor dust measurements, dry room window sill dust measurements, and the entrance dust measurements.

Table 6-14 Parameter estimates for the dry room window sill dust regressions

		Observed Variables	Approx. principle components	Interior versus Exterior
Paint lead (X)	Dry rooms	0.04±0.14		
	Wet rooms	0.13±0.14		
	Exterior	-0.06±0.09		-0.07±0.09
	Average		0.11±0.14	
	Wet - dry		0.00±0.12	0.00±0.12
	Int - Ext		0.12±0.10	
	Interior			0.16±0.14
Proportion damaged paint (D)	Dry rooms	-6.8±15.1		
	Wet rooms	-2.1±6.4		
	Exterior	2.9±4.4		1.8±4.4
	Average		0.9±10.4	
	Wet - dry		2.5±7.3	0.6±8.8
	Int - Ext		-1.4±5.1	
	Interior			-5.0±13.4
Painted surface area (A)	Dry rooms	-0.29±0.30		
	Wet rooms	-0.10±0.31		
	Exterior	-0.01±0.21		-0.04±0.21
	Average		-0.54±0.46	
	Wet - dry		0.10±0.24	0.12±0.24
	Int - Ext		-0.13±0.20	
	Interior			-0.46±0.41
Soil (S)	Entrance	0.18±0.26		
	Drip line	0.31±0.24		
	Remote	-0.17±0.27		-0.12±0.27
	Average		0.37±0.31	
	Entrace-drip line		-0.05±0.20	-0.07±0.20
	Close-remote		0.24±0.24	
	Close			0.48±0.31
Other terms	# of wet rooms	-0.63±0.91	-0.23±0.87	-0.42±0.89
	# of dry rooms	0.45±0.76	0.39±0.75	0.43±0.76
	Local traffic (Traf)	0.09±0.19	0.07±0.19	0.04±0.19
	Unit age (Age)	0.37±0.88	0.41±0.92	0.42±0.89
	Number of Occupants	0.26±0.52	0.25±0.53	0.32±0.53
Significant interactions and quadratic terms		Wrooms^2=2.98 Traf*age=0.41	Wrooms^2=2.76	Wrooms^2=3.10
Floor level (p-value)		0.8068	0.8137	0.8523
Adjusted R-square		0.255	0.227	0.240
Number of dwelling units, n		161	161	161

Table 6-15 Parameter estimates for the wet room window sill dust regressions

		Observed Variables	Approx. principle components	Interior versus Exterior
Paint lead (X)	Dry rooms	0.01±0.21		
	Wet rooms	0.28±0.21		
	Exterior	-0.01±0.13		-0.04±0.14
	Average		0.37±0.22	
	Wet - dry		0.15±0.25	0.14±0.25
	Int - Ext		0.12±0.15	
	Interior			0.33±0.22
Proportion damaged paint (D)	Dry rooms	16.8±22.9		
	Wet rooms	-0.4±7.7		
	Exterior	1.8±5.4		2.9±5.2
	Average		7.5±14.9	
	Wet - dry		-4.4±8.6	-6.9±11.3
	Int - Ext		-0.7±6.3	
	Interior			7.7±21.5
Painted surface area (A)	Dry rooms	-0.14±0.44		
	Wet rooms	-0.23±0.44		
	Exterior	0.40±0.30		0.40±0.31
	Average		0.10±0.67	
	Wet - dry		0.01±0.33	-0.01±0.34
	Int - Ext		-0.37±0.29	
	Interior			-0.34±0.60
Soil (S)	Entrance	0.31±0.39		
	Drip line	-0.24±0.36		
	Remote	-0.11±0.39		-0.08±0.39
	Average		0.00±0.43	
	Entrace-drip line		0.31±0.33	0.30±0.33
	Close-remote		0.05±0.34	
	Close			0.07±0.42
Other terms	# of wet rooms	-0.25±1.24	-0.56±1.36	-0.37±1.25
	# of dry rooms	-0.54±1.03	-0.45±1.02	-0.50±1.04
	Local traffic (Traf)	0.25±0.27	0.24±0.28	0.27±0.28
	Unit age (Age)	-0.02±1.47	-0.22±1.51	-0.17±1.49
	Number of Occupants	0.28±0.77	0.32±0.78	0.24±0.80
Significant interactions and quadratic terms		Xwet^2=0.05	Xavg^2=0.07	Xint^2=0.06
Floor level (p-value)		0.0246	0.0156	0.0184
Adjusted R-square		0.304	0.287	0.288
Number of dwelling units, n		106	106	106

Table 6-16 Parameter estimates for the dry room window well dust regressions

		Observed Variables	Approx. principle components	Interior versus Exterior
Paint lead (X)	Dry rooms	0.06±0.28		
	Wet rooms	0.06±0.30		
	Exterior	0.07±0.23		0.04±0.25
	Average		-0.20±0.55	
	Wet - dry		-0.01±0.24	-0.02±0.23
	Int - Ext		0.09±0.19	
	Interior			0.09±0.25
Proportion damaged paint (D)	Dry rooms	16.8±48.5		
	Wet rooms	-5.4±21.6		
	Exterior	0.4±9.7		0.5±9.1
	Average		-3.3±27.4	
	Wet - dry		-3.4±21.8	-9.8±25.9
	Int - Ext		-4.1±15.7	
	Interior			8.9±45.6
Painted surface area (A)	Dry rooms	-0.07±0.81		
	Wet rooms	0.06±0.72		
	Exterior	0.17±0.48		0.20±0.50
	Average		0.53±1.24	
	Wet - dry		-0.03±0.64	-0.01±0.62
	Int - Ext		-0.03±0.46	
	Interior			0.18±1.05
Soil (S)	Entrance	-0.17±0.63		
	Drip line	-0.35±0.51		
	Remote	0.10±0.58		0.15±0.59
	Average		-0.17±0.77	
	Entrace-drip line		0.03±0.48	0.08±0.45
	Close-remote		-0.30±0.53	
	Close			-0.58±0.71
Other terms	# of wet rooms	0.05±1.88	-0.84±2.11	-0.07±1.92
	# of dry rooms	-1.12±1.57	-0.73±1.54	-1.20±1.54
	Local traffic (Traf)	0.26±0.53	-0.13±0.67	0.21±0.62
	Unit age (Age)	2.56±2.23	3.16±2.44	2.46±2.25
	Number of Occupants	0.55±0.84	0.27±0.89	0.52±0.88
Significant interactions and quadratic terms				
Floor level (p-value)		0.0333	0.0238	0.0451
Adjusted R-square		0.244	0.175	0.224
Number of dwelling units, n		63	63	63

Table 6-17 Parameter estimates for the wet room window well dust regressions

		Observed Variables	Approx. principle components	Interior versus Exterior
Paint lead (X)	Dry rooms	-0.07±0.36		
	Wet rooms	0.24±0.30		
	Exterior	-0.46±0.39		-0.45±0.38
	Average		0.14±0.58	
	Wet - dry		0.16±0.31	0.18±0.29
	Int - Ext		0.15±0.29	
	Interior			0.15±0.38
Proportion damaged paint (D)	Dry rooms	-48.8±107.2		
	Wet rooms	14.0±41.4		
	Exterior	13.9±19.8		14.4±17.9
	Average		-24.5±78.2	
	Wet - dry		11.1±74.0	30.3±67.1
	Int - Ext		-8.7±39.2	
	Interior			-31.1±78.4
Painted surface area (A)	Dry rooms	-0.74±1.08		
	Wet rooms	-0.03±1.15		
	Exterior	0.29±0.51		0.30±0.52
	Average		-0.46±1.34	
	Wet - dry		0.18±1.02	0.34±0.90
	Int - Ext		-0.52±0.59	
	Interior			-0.79±1.16
Soil (S)	Entrance	-0.56±0.99		
	Drip line	0.19±0.77		
	Remote	0.20±0.72		0.15±0.78
	Average		0.10±1.04	
	Entrace-drip line		-0.23±0.94	-0.28±0.83
	Close-remote		-0.24±0.64	
	Close			-0.37±0.78
Other terms	# of wet rooms	1.02±1.93	0.95±2.09	1.03±1.88
	# of dry rooms	-0.24±1.58	0.05±1.70	-0.27±1.55
	Local traffic (Traf)	0.11±0.57	0.10±0.64	0.19±0.58
	Unit age (Age)	0.91±2.66	1.03±2.85	0.97±2.64
	Number of Occupants	0.11±1.41	-0.37±1.72	0.11±1.55
Significant interactions and quadratic terms				
Floor level (p-value)		0.0599	0.1886	0.0571
Adjusted R-square		0.375	0.260	0.386
Number of dwelling units, n		54	54	54

Table 6-18 Parameters, with 95% confidence interval, for two possible models for identifying sources of lead in dry room floor dust

Variable		Model	
		Dust lead concentration	Dust lead loading with dust loading
Paint lead loading	dry room	0.07±0.10	0.07±0.09
	wet room	0.04±0.10	0.07±0.09
	exterior	-0.02±0.06	0.00±0.06
Proportion damaged paint	dry room	4.76±16.45	10.45±14.93
	wet room	0.51±4.78	0.55±4.32
	exterior	1.34±3.74	1.47±3.40
Painted surface area	dry room	-0.10±0.22	-0.13±0.20
	wet room	0.06±0.23	-0.01±0.21
	exterior	0.02±0.15	0.03±0.14
Soil lead concentration	entrance	0.13±0.19	0.16±0.18
	drip line	0.12±0.18	0.07±0.17
	remote	-0.09±0.20	-0.07±0.19
Number of wet rooms		0.44±0.59	0.30±0.54
Number of dry rooms		0.49±0.54	0.24±0.50
Local traffic volume		0.08±0.14	0.11±0.13
dwelling unit age		0.48±0.64	0.25±0.58
Number of occupants		-0.13±0.38	-0.07±0.34
Dust weight per unit area			0.67±0.10

Table 6-19 Parameters, with 95% confidence interval, for two possible models for identifying sources of lead in dry room window sill dust

Variable		Model	
		Dust lead concentration	Dust lead loading with dust loading
Paint lead loading	dry room	0.04±0.14	0.05±0.13
	wet room	0.13±0.14	0.12±0.13
	exterior	-0.06±0.09	-0.04±0.09
Proportion damaged paint	dry room	-6.78±15.07	-7.25±13.94
	wet room	-2.12±6.43	-1.09±5.97
	exterior	2.89±4.43	3.27±4.10
Painted surface area	dry room	-0.29±0.30	-0.36±0.28
	wet room	-0.10±0.31	-0.06±0.29
	exterior	-0.01±0.21	0.01±0.19
Soil lead concentration	entrance	0.18±0.26	0.17±0.24
	drip line	0.31±0.24	0.26±0.22
	remote	-0.17±0.27	-0.20±0.25
Number of wet rooms		-0.63±0.91	-0.41±0.84
Number of dry rooms		0.45±0.76	0.06±0.72
Local traffic volume		0.09±0.19	0.16±0.18
dwelling unit age		0.37±0.88	0.51±0.82
Number of occupants		0.26±0.52	0.24±0.49
Dust weight per unit area			0.73±0.11

Table 6-20 Parameters, with 95% confidence interval, for two possible models for identifying sources of lead in entrance floor dust

Variable		Model	
		Dust lead concentration	Dust lead loading with dust loading
Paint lead loading	dry room	-0.05±0.08	-0.02±0.07
	wet room	0.04±0.08	0.04±0.07
	exterior	-0.04±0.05	-0.02±0.05
Proportion damaged paint	dry room	1.23±8.84	2.26±7.70
	wet room	-0.23±3.67	-0.53±3.20
	exterior	-1.59±2.46	-0.41±2.16
Painted surface area	dry room	-0.06±0.19	-0.09±0.16
	wet room	0.11±0.18	0.02±0.15
	exterior	-0.03±0.12	0.00±0.10
Soil lead concentration	entrance	0.30±0.15	0.29±0.13
	drip line	0.11±0.14	0.04±0.13
	remote	-0.07±0.16	-0.03±0.14
Number of wet rooms		0.15±0.47	0.14±0.41
Number of dry rooms		0.20±0.42	-0.04±0.37
Local traffic volume		-0.02±0.11	0.02±0.10
dwelling unit age		0.23±0.51	0.30±0.44
Number of occupants		-0.04±0.29	-0.02±0.26
Dust weight per unit area			0.72±0.07

ERRATA SHEET

Corrections to the published report:

Data Analysis of Lead on Soil and Dust
EPA 747-R-93-011, September 1993

November 9, 1995

1) One value in the report, which was cited several times, is incorrect (it should be 5.2 instead of 2.7). The sentences containing the incorrect value are listed below, along with the corrected sentences. In one case other numbers in the same sentence change as a result of the correction.

Page xvii in the Executive Summary, second paragraph:

As printed: While the measurement error is relatively large (about 95 percent of soil lead measurements will be within a factor of 2.7 of the true concentration), it is small compared to the differences in soil lead concentrations between locations.

Corrected: While the measurement error is relatively large (about 95 percent of soil lead merasurements will be within a factor of 5.2 of the true concentration), it is small compared to the differences in soil lead concentrations between locations.

Page 33, last paragraph:

As printed: Assuming the soil lead measurement errors have a log normal distribution, we would expect 95 percent of the soil lead concentrations to be within a factor of 2.7 of the true average lead concentration in the area of interest. For example, if a soil lead measurement is 47 ppm, the true lead concentration in the area of interest around the soil sample is most likely between 17 and 127 ppm (47/ 2.7 and 47*2.7 respectively).

Corrected: Assuming the soil lead measurement errors have a log normal distribution, we would expect 95 percent of the soil lead concentrations to be within a factor of 5.2 of the true average lead concentration in the area of interest. For example, if a soil lead measurement is 47 ppm, the true lead concentration in the area of interest around the soil sample is most likely between 9 and 244 ppm (47/ 5.2 and 47*5.2 respectively).

Page 35, first paragraph:

As printed: Although the measurement error might be considered to be large (i.e., within a factor of 2.7), the measurement error is small compared to the differences in soil lead concentrations between homes of different ages.

Corrected: Although the measurement error might be considered to be large (i.e., within a factor of 5.2), the measurement error is small compared to the differences in soil lead concentrations between homes of different ages.

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. EPA 747-R-93-011	2.	3. Recipient's Accession No.

4. Title and Subtitle

Data Analysis of Lead in Soil and Dust

5. Report Date
September, 1993

6.

7. Author(s)
Westat, Inc.

8. Performing Organization Rept. No.

9. Performing Organization Name and Address

Westat, Inc.
1650 Research Blvd.
Rockville, MD 20850

10. Project/Task/Work Unit No.

11. Contract (C) or Grant (G) No.
(C) 68-D9-0174
(C) 68-D2-0139
(C) 68-D3-0011

12. Sponsoring Organization Name and Address
U.S. Environmental Protection Agency
Office of Pollution Prevention and Toxics
Washington, DC 20460

13. Type of Report & Period Covered
Technical Report

14.

15. Supplementary Notes

16. Abstract (Limit: 200 words)

In the National Survey of Lead-Based Paint in Housing, conducted by the EPA and HUD, samples of soil interior dust were collected for lead analysis. Paint lead loadings were measured using a portable XRF device. This report (1) presents a detailed analysis of the soil and dust data from private dwelling units, (2) uses regression analysis to assess the statistical association between soil lead, paint lead, and dust lead, and discusses the impact of measurement error on the classification of dwelling units with high lead levels. No outliers were found among the soil lead measurements. About one percent of the dust lead concentrations are clearly outliers and were removed for subsequent analysis. The dust, soil, and paint lead concentrations all increase as the age of the dwelling unit increases. Regression was used to identify possible sources of the lead in soil and dust. Although the regression results explain only a small portion of the variance, significant relationships were found. In particular, both soil and paint lead appear to contribute to interior dust lead, with damaged exterior paint contributing to the dust lead in the dwelling unit entrance. Both lead paint and auto emissions appear to contribute to soil lead.

17. Document Analysis a. Descriptors

Environmental Contaminants

b. Identifiers/Open-Ended Terms

Soil lead, dust lead, lead-based paint, regression analysis, National Survey of Lead-Based Paint in Housing

c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Available to the public from NTIS, Springfield, VA	Unclassified	125
	20. Security Class (This Page) Unclassified	22. Price

(See ANSI-Z39.18) *See Instructions on Reverse*

OPTIONAL FORM 272 (4-77)
(Formerly NTIS-35)
Department of Commerce