United States
Environmental
Protection Agency

Office of
Pollution Prevention
and Toxics

EPA 747-R96-002
May 1996

EPA

Distribution of Soil Lead in the Nation's Housing Stock

Distributions of Soil Lead in the Nation's Housing Stock

This work was conducted under contract
number 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

May, 1996

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.

Table of Contents

Tables

Tables (continued)

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 Cindy Stroup, Samuel F. Brown, Brad Schultz, Philip E. Robinson, and Sineta Wooten for their guidance and support throughout this research.

Executive Summary

The primary objective of this study was to supplement the prior reports on the National Survey of Lead-Based Paint in Housing through additional data analyses specifically focusing on the relationship between lead in exterior soil (a potential source of lead hazard in homes) and housing unit characteristics. The 1987 amendments to the Lead-Based Paint Poisoning Prevention Act required the Secretary of Housing and Urban Development (HUD) to "estimate 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 act, HUD initiated and conducted the National Survey of Lead-Based Paint in Housing, or the National Survey in 1990. The survey results were published in the Environmental Protection Agency's (EPA) *Report On The National Survey of Lead-Based Paint In Housing* document the National Survey and presented data on the extent and characteristics of lead hazards in homes.

The National Survey inspected 381 housing units (284 privately-owned and 97 public) for lead in paint on interior and exterior surfaces, lead in interior dust, and lead in exterior soil. The study population was designed to be representative 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 lead-based paint to consumers and the use of such paint in residences. The National Survey was conducted between December 1989 and March 1990 in 30 counties across the 48 contiguous states. These counties were selected to represent both the public and privately-owned housing stock across the 48 contiguous states.

The purpose of this report is to supplement discussions on soil lead prevalence in the prior reports on the National Survey by presenting findings on the prevalence and concentrations of lead in soil around private and public housing units in the United States. These findings included estimates of the number of housing units with different soil lead concentrations, nationally, by building age, by Census region, and by degree of urbanization; and summaries of the statistical associations between soil lead concentrations and soil location, building age, degree of urbanization, Census region, and the presence and condition of interior and exterior lead-based paint.

The quality of the private and public housing data was statistically evaluated to determine the suitability of the soil lead data for the analyses needed in this study. The privately-owned homes sampled in the National Survey were judged to be representative of the private housing stock nationally. Therefore, the descriptive statistics presented in the private housing data tables and the results from the analyses on the private housing data can be viewed as applicable to private housing nationally and useful in policy analysis and decision making. In contrast, the sampled public housing units were not considered representative of the public housing stock nationally, and the impact of the large amount of missing soil data (70%) on the tables and analysis results was expected to be significant. The public housing data tables and results from the analyses on public housing should therefore be viewed only as descriptive of those samples collected.

Under Section 403 of Title X, EPA has established health-based interim standards for soil lead concentrations and action recommendations for each standard. The agency recommends that "interim controls to change use patterns and establish barriers" should be implemented for areas that are expected to be used by children where soil lead concentrations are between 400 and 5,000 parts per million (ppm). Within this range, the degree of activity should be "commensurate with the expected risk posed by the bare soil considering both the severity of [lead] exposure...and the likelihood of the children's exposure." For areas where contact by children is less likely or less frequent, the "interim controls" should be implemented when soil lead concentrations are between 2,000 and 5,000 ppm. Moreover, the agency recommends the "abatement of soil" with lead concentrations above 5,000 ppm regardless of the likelihood of children's exposure.

Using the data from the National Survey, it is estimated that 23 percent, or 18 million, of the privately-owned homes in the United States built before 1980 have soil lead levels that exceed the 400 ppm "interim control" guideline. An estimated 8 percent, or 6 million, of the privately-owned homes in the United States built before 1980 have soil lead levels that exceed the 2,000 ppm "interim control" guideline. Finally, an estimated 3 percent, or 2.5 million, of the privately-owned homes in the United States built before 1980 have soil lead levels that exceed the 5,000 ppm soil abatement guideline. The prevalence and distribution of soil lead concentrations in public housing was not estimated due to the considerable number of public housing units in the National Survey for which no soil was available for sampling.

This study assessed the associations between the soil lead concentrations at different locations and the presence and condition of interior and exterior lead-based paint to determine which characteristics and factors specific to the housing unit are good predictors of soil lead. Additional variables also considered to be related to soil lead included the average daily traffic flow in the neighborhood of the housing unit (for private housing only) and the number of family units in the development (for public housing only), both of which were used to estimate the impact of the housing unit's environment on soil lead.

Private Housing

The strongest statistical predictor of soil lead was found to be the building age. Building age measures the length of time since the construction of the building and, in many cases, may be the last major disturbance of soil. For private housing units, soil lead around homes built before 1940 were significantly greater than lead in soil around homes built between 1960 and 1979. Similarly, soil lead around public housing units built before 1950 are significantly greater than lead in soil around homes built between 1960 and 1979.

The Census region (Northeast, Midwest, South, West) in which the housing unit was located was also an important predictor of soil lead levels. The data analysis showed that after adjusting for the age of the housing unit, soil around private housing units in the Northeast region has, on average, higher lead concentrations than in any other region, and soil in the Midwest region has on average, higher lead concentrations than those in either the West or South regions. One possible explanation is that the Northeast and Midwest are more industrialized, e.g., have the highest level of industrial productivity, of the four regions of the United States.

Another finding was soil lead levels around homes in urban, suburban, and rural areas were unexpectedly not significantly different, after adjusting for building age and other factors. Explanations of this result include one or more of the following: the distribution of privately-owned homes where soil lead measurements were not taken corresponds to sites which were expected to have high soil lead concentrations (33 of the 93 sampled private housing units in large metropolitan areas have at least one missing soil lead measurement), the correlations between the degree of urbanization and other factors, such as traffic, might be reducing the effect of highly urbanized areas, and the random variation in the data associated with the selection of the homes.

After adjusting for building age, Census region, and other factors, the presence of lead-based paint was an important predictor of soil lead at all three locations. The condition of lead-based paint, however, was not an important predictor of soil lead at any of the three soil locations.

Public Housing

Soil lead samples were available for only 30 percent (29 of 97) of the sampled public housing units, and the distribution of public housing units with soil lead samples was not consistent with national distributions. These problems prevented any reliable national estimates of soil lead prevalence in public housing from being calculated.

Although no estimates for the effects of the degree of urbanization could be made with respect to public housing developments, the relationship between soil lead and housing unit characteristics in public housing was analyzed with respect to building age and the presence and condition of lead-based paint. The findings showed that these relationships were similar to those in private housing data. The building age was the most important predictor of soil lead concentrations. The Census region in which the development was located was an important predictor of soil lead after adjusting for the age of the development. Housing unit variables that were correlated with soil lead but were not significant predictors of soil lead after adjusting for the age of the development and the Census region included the number of family units in the public housing development (which was slightly correlated with the development's building age) and the condition of lead-based paint in and around the housing unit.

1. Introduction

The 1987 amendments to the Lead-Based Paint Poisoning Prevention Act required the Secretary of Housing and Urban Development (HUD) to "estimate 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 act, HUD initiated the National Survey of Lead-Based Paint in Housing, or the National Survey which was completed in 1990. The National Survey produced a detailed, statistically valid, national database on the extent of lead-based paint and lead in soil and dust. These data have been and continue to be analyzed to support the development of Federal policy and programs with respect to the lead hazard in homes.[1]

Issues currently before the U.S. Environmental Protection Agency involve the relationships between housing unit characteristics and lead exposure levels. 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. In addition, lead-based paint, primarily exterior lead-based paint, is believed to be a significant contributor to soil lead contamination. Although the National Survey did not collect data on direct measures of lead exposure, such as children's blood lead levels, an analysis of the relationship between soil lead and housing unit characteristics may aid in understanding the relationship between housing unit characteristics and potential lead exposure.

EPA is developing health-based standards for dust, paint, and soil lead concentrations under Section 403 of the Residential Lead-Based Paint Hazard Reduction Act of 1992 (Title X). These standards are published as EPA's *Guidance on Identification of Lead-Based Paint Hazards*[2] and referred to as the 403 Interim Final Rule.

1.1 Purpose of Report

The purpose of this report is to supplement the prior reports on the National Survey by addressing the following objectives:

- Present findings from the National Survey on the prevalence and concentrations of lead in soil around private and public housing units in the United States, including estimates of the number of housing units with different soil lead concentrations, nationally, by building age, Census region, and degree of urbanization;

- Summarize the statistical associations between soil lead concentrations and soil location, building age, degree of urbanization, Census region, and the presence and condition of interior and exterior lead-based paint;

---

[1] A complete discussion of the National Survey, including the design, sample collection protocol, and results from the data analyses, can be found in EPA's *Report on the National Survey of Lead-Based Paint in Housing*.

[2] *Guidance on Identification of Lead Based Paint Hazards*, Federal Register, v-60 (175): September 11, 1995.

1.2 Overview of the National Survey

The National Survey was conducted by HUD. In that sample survey, 381 housing units, 284 private and 97 public, were inspected for lead in paint on interior and exterior surfaces, lead in interior dust, and lead in exterior soil. The objective of the National Survey was to obtain data for estimating the following:

- The number of housing units with lead-based paint;
- The surface area of lead-based paint in housing, used to develop an estimate of national abatement costs;
- The condition of the paint;
- The prevalence of lead in house dust and in soil around the perimeter of residential structures; and
- The characteristics associated with varying levels of potential lead hazards in housing in order 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 lead-based paint to consumers and the use of such paint in residences. The survey was conducted between December 1989 and March 1990 in 30 counties across the 48 contiguous states.

The 30 counties were randomly selected from the approximately 3,000 counties in the United States to represent the nation's private and public housing stock built before 1980. The counties were stratified by Census region (Northeast, South, Midwest, and West) and climate (mild or severe weather) and selected with probability proportion to size. The private housing units were selected as follows. Within each sampled county, five census blocks were randomly selected and a list of every housing unit within each census block was developed. An initial sample of the listed units was randomly selected for in-person screening visits to establish eligibility. An average of 20 housing units per census block were screened and an average of 11 were found to be eligible. From the eligible housing units, two (plus backups) were randomly selected.

The public housing units were selected as follows. Within each sampled county, lists of the Public Housing Authority (PHA) housing developments, including the numbers and types of units in the development, were created from lists supplied by HUD. The lists for each of the 30 counties were merged, sorted by the age of the development, and a stratified random sample of 110 developments was drawn. Within each of the selected developments, one unit was randomly selected.

Within each sampled private and public housing unit, two rooms were randomly selected for inspection -- one with plumbing, a "wet room," and the other without plumbing, a "dry room." In each room, field technicians inventoried painted surfaces, measured the surface area, and assessed the condition of the paint. They also measured the lead loadings on randomly selected painted surfaces with portable X-ray fluorescence (XRF) analyzers. 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.

Exterior soil samples and interior dust samples were also collected. Generally, three soil core samples were taken from each dwelling unit: one outside the main entrance to the building, a second along the drip line (soil next to the housing unit), and a third at a remote location away from the building but still on the property. The drip line and the remote samples were usually collected on the same, randomly selected side of the house as the exterior XRF paint lead measurement. Dust samples were collected on floors, window wells, and windows sills in the wet and dry rooms and from the floor immediately inside the main entrance to the dwelling unit. Dust samples were also collected from common areas inside private multifamily and public housing units. Since the sample size for the common area dust samples was small, they are not discussed in this report. Both dust and soil samples were sent to laboratories for lead analysis.

Midwest Research Institute (MRI) was the subcontracting laboratory responsible for the analysis of both soil and dust samples. MRI and its subcontractor, Core Laboratories, with a site in Casper, Wyoming and another in Aurora, Colorado, analyzed the samples for lead. The Casper facility analyzed both soil and dust samples, while the Aurora facility analyzed only dust samples. A total of 3,231 samples, 1,053 soil samples and 2,178 dust samples, were analyzed. The dust samples were analyzed by graphite furnace atomic absorption (GFAA) spectroscopy. The soil samples were analyzed by inductively coupled plasma-atomic emission spectrometry (ICP-AES). Internal checks, including duplicate injections to measure instrument precision, and external checks, including the analysis of split samples to measure the variability from sample handling prior to analysis, were used to track performance. In addition, performance check samples were analyzed to measure the accuracy of the analytical procedures. The results on the internal, external, and performance checks were satisfactory, meeting most of the data quality objectives. MRI's *Analysis of Soil and Dust Samples for Lead (Pb), Final Report*[3] details its methodology and data quality procedures.

---

[3] *Analysis of Soil and Dust Samples for Lead (Pb), Final Report*, May 8, 1991. Prepared under contract to the U.S. Environmental Protection Agency. EPA Contract No. 68-02-4252.

2. Conclusions

This chapter presents the overall conclusions from the analyses of possible predictors of lead in soil. The specific objectives and analytic requirements of many of these analyses were not foreseen when the National Survey was designed and implemented. Therefore, the suitability of the data for analysis, which includes a review of what the data actually represent were evaluated. The conclusions about the suitability of the data for analysis and results from the analyses are presented followed by a more detailed explanation of the conclusion.

1. **The private housing data in the National Survey can be viewed as representative of the nation's housing stock and suitable for the analysis.**

For private housing units, the distribution of households in the National Survey was not significantly different from the distribution of households in the American Housing Survey with respect to building age. Differences with respect to the Census region, though, were only marginally significant in that more dwelling units located in the South were sampled in the National Survey than expected based on the American Housing Survey. Additionally, soil samples were taken at 94 percent of the private housing units in the National Survey. Because the distributions of households in the National Survey were not significantly different from those found in American Housing Survey, only a small percentage (six percent) of the sampled privately-owned homes had no soil lead measurements, and a large amount of data was available (over 250 observations for each model), there are no apparent reasons why inferences cannot be drawn from analyses for private homes.

2. **The public housing data in the National Survey can not be viewed as representative of the nation's public housing stock and results about public housing should be viewed with caution.**

For public housing units, differences between the distribution of sampled public housing units and the distribution of all public housing units, provided by HUD, are significant based on both Census region and building age. Moreover, problems with the lack of soil lead measurements make analyses of the data difficult to interpret. Soil samples were available at only 30 percent (29 of 97) of the sampled public housing developments. Given both the distributional inequality and the relatively small number of public housing units where soil samples were taken (n=29), all conclusions about public housing units and results from analyses of the public housing data should be viewed with caution.

3. **The strongest statistical predictor of soil lead in private and public housing for all sample locations is the housing unit's date of construction.**

The date of construction, or building age, measures the amount of time since the construction of the building and, in many cases, is the last major disturbance of soil. Thus, the building age likely measures the length of time lead -- from the housing unit and/or neighboring activity sources -- has been accumulating on the soil. For private housing units, soil lead around homes built before 1940 were significantly greater than lead in soil around homes built between 1960 and 1979. Similarly, soil lead around public housing units built before 1950 are significantly greater than lead in soil around homes built between 1960 and 1979.

4. **Additional significant predictors of soil lead in private housing include the Census region, the interaction between the building age and the Census region, the presence of lead-based paint, and the average daily traffic flow.**

After adjusting for the housing unit's age, soil around privately-owned homes in the Northeast region was estimated to have, on average, higher lead concentrations than in any other region. In addition, soil around privately-owned homes in the Midwest region was estimated to have, on average, higher lead concentrations than in either the West or South regions. Soil lead concentrations at the remote location around privately-owned homes in the Midwest region built between 1940 and 1949, however, were estimated to have lower soil lead concentrations than in any other region. One possible explanation for the average higher soil lead concentrations in the Northeast and Midwest regions is that these regions are the most industrialized, e.g., have the highest level of industrial productivity, of the four regions of the Unites States.

The presence of lead-based paint was shown to have a significantly positive effect on soil lead concentrations at all three locations, but to a larger extent at the drip line and entryway locations. In addition, the traffic flow (a source of lead from automobile emissions) in the neighborhood around the private housing unit was shown to have a significantly positive effect on soil lead concentrations at the remote location. These results support the concerns in the 403 Interim Final Rule about lead in residential soil from "lead-based paint and...as the result of point source emissions or leaded gasoline."

5. **The degree of urbanization and condition of lead-based paint are not significant predictors of lead in soil in private housing.**

Soil lead levels around homes in urban, suburban, and rural areas were unexpectedly not significantly different after adjusting for other factors such as building age, Census region, and the presence of lead-based paint. Explanations of this result include are likely to include one or more of the following: the distribution of the missing soil lead measurements corresponds to sites which were expected to have high soil lead concentrations (33 of the 93 sampled private housing units in large metropolitan areas have at least one missing soil lead measurement); the correlations between the degree of urbanization and other factors, such as building age or traffic, might be reducing the significance of the effect of highly urbanized areas; and the random variation in the data associated with the selection of the homes.

After adjusting for the housing unit's building age, Census region, and presence of lead-based paint, the effect of the condition of lead-based paint on soil lead levels was also unexpectedly insignificant. This result is likely due to the fact that the condition of lead-based paint is correlated with the building age, the Census region, and the presence of lead-based paint and does not explain any significant variation in the soil lead levels after adjusting for the building age, Census region, and presence of lead-based paint.

6. **The only other significant predictors of soil lead in public housing is the Census region.**

After adjusting for the building age, soil around public housing developments in the Midwest and West regions was estimated to have, on average, higher lead concentrations than in the South region. No estimates of soil lead prevalence around public housing developments could be made for the Northeast region because only one sampled public housing development had soil samples. In addition, the effect of the degree of urbanization could not be analyzed because no such data were collected. The condition of lead-based paint and the number of family units, both positively correlated with soil lead, were not significant predictors of lead-based paint after adjusting for building age and Census region.

7. **The results for the private housing data can be viewed as applicable to private housing nationally and useful in policy analysis and decision making, but the results for the public housing data would be viewed only as descriptive of those housing units sampled.**

The quality of the private and public housing data was statistically evaluated to determine the suitability of the soil lead data for the analyses needed in this study. The privately-owned homes sampled in the National Survey were judged to be representative of the private housing stock nationally. Therefore, the descriptive statistics presented in the tables for and the results from the analyses on the private housing data can be viewed as applicable to private housing nationally and useful in policy analysis and decision making. In contrast, the sampled public housing units were not considered representative of the public housing stock nationally, and the impact of the large amount of missing soil data (70%) on the tables and analysis results was expected to be significant. The tables and analysis results for public housing should therefore be viewed only as descriptive of those samples collected.

3. Descriptive Soil Lead and Housing Unit Statistics

This chapter discusses the soil lead data; housing unit characteristics, including the representativeness of the sampled housing units; and soil lead prevalence levels in private and public housing units. It also presents summaries of the soil lead and housing unit characteristic data in tabular form. Sample weights were used in the estimates displayed in most of the tables. This was done so that inferences could be drawn from these estimates about the populations of private and public housing. The estimates presented in these tables are, under certain circumstances that are discussed and evaluated in this chapter, representative of private and public housing nationally. The information presented here is used as background information for the data analyses presented in Chapters 4 and 5.

3.1 Soil Lead Data

The sampling protocols required that soil be collected from three locations around each sampled dwelling unit. Soil samples were to be taken outside the main entrance to the building, at a selected location along the drip line of an exterior wall, and at a remote location (away from the building, but still on the property). The field and laboratory protocols for sampling and analysis are presented briefly in Chapter 1, in *Data Analysis of Lead in Soil and Dust*,[4] and in MRI's *Analysis of Soil and Dust Samples for Lead (Pb): Final Report*.[5]

Basic weighted descriptive statistics for private and public housing units are presented in Tables 1 and 2. These statistics include the sample mean, standard deviation, coefficient of variation, selected percentiles, geometric mean, and geometric standard deviation of the soil lead measurements for the entrance, drip line, and remote soil lead measurements.[6] 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. It is useful for describing data such as soil lead concentration data that are always greater then equal to zero. The geometric mean and standard deviation are often used for right skewed data, because they reduce the impact of extremely large measurements.

Private Housing Data

In some cases, such as around urban private housing units with all areas around the housing unit paved or with no soil on the property, soil samples were not taken. Of the 284 private housing units in the National Survey sample, 18 housing units had no soil samples taken and another 26 housing units were missing data from one or two of the three soil locations. Thus, a total of 44 housing units were missing one or more soil samples. Of the 18 housing units without soil data, 14 were located in large metropolitan urban areas, 15 were in the Northeast Census region, and 12 were built before 1940. Of the 44 housing units with some missing soil data, 33 were located in large metropolitan urban areas, 21 were

---

[4] *Data Analysis of Lead in Soil and Dust*, September, 1993. EPA Report number 747-R-93-011.

[5] *Analysis of Soil and Dust Samples for Lead (Pb), Final Report*, May 8, 1991. Prepared under contract to the U.S. Environmental Protection Agency. EPA Contract No. 68-02-4252.

[6] Additional analyses of the soil lead data may be found in the following reports: HUD's *Comprehensive and Workable Plan for the Abatement of Lead-Based Paint in Privately Owned Housing: Report to Congress*, and EPA's *Data Analysis of Lead in Soil and Dust* and *Report on the National Survey of Lead-Based Paint*

Table 1. Descriptive statistics (weighted) for the lead measurements in soil samples at each soil location in private housing units

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 1%	6,829	9,965	2,974
upper 5%	1,377	1,447	603
upper decile	775	860	278
upper quartile	225	234	120
median	64.8	56.2	46.7
lower quartile	28.9	21.6	18.5
minimum	2.84	1.16	1.45
Geometric mean (ppm)	85	74	46
Geometric standard deviation (ppm)	2.11	1.80	1.81

Table 2. **Descriptive statistics (weighted) for the lead measurements in soil samples at each soil location in public housing units**

Set of data	Entrance samples	Drip line samples	Remote samples
Number of measurements	26	28	29
Arithmetic mean (ppm)	127	117	83
Percentiles (ppm)			
maximum	527	871	614
upper 1%	527	871	614
upper 5%	483	871	243
upper decile	438	265	209
upper quartile	186	140	99.5
median	44.0	31.2	42.9
lower quartile	23.1	22.0	23.1
minimum	8.10	10.6	5.67
Geometric mean (ppm)	55	55	44
Geometric standard deviation (ppm)	1.27	1.28	1.19

in the Northeast Census region, and 20 were built before 1940. A more detailed distribution of the missing data, including totals for private homes in the National Survey, can be found in Table 3.

Only 24 out of 762, or 3 percent, of the soil lead concentration measurements were reported below the method detection limit, which ranged from 3 to 20 ppm.[7] A common practice of replacing the measurements below the detection limit with one-half of the detection limit was followed. The replaced values were consistent with the distribution of all soil lead measurements. Accordingly, the handling of the measurements below the detection limit is expected to have no significant effect on the statistical analysis results.

Public Housing Data

As with private housing, soil samples were not collected around all of the sampled public housing units. Unlike the private housing data, where soil samples were taken at all but 6 percent of the homes, more than 70 percent of the public housing units had no soil samples taken. This considerably larger percentage of missing data has the potential to significantly bias the results of any analysis. Of the 97 public housing units in the National Survey sample, 68 had no soil samples, and an additional four housing units were missing data from one or two of the three soil locations. A more detailed distribution of the missing data for public housing units in the National Survey can be found in Table 4. No soil lead concentrations for public housing units that were sampled were below the instrument detection limit.

3.2 Soil Lead Prevalence

The weighted sample geometric mean soil lead concentrations at the drip line, entryway, and remote locations are 74, 85, and 46 ppm, respectively, for private homes and 55, 55, and 44, respectively, for public housing units. Paired differences between the log-transformed measurements were used to determine if the differences in weighted geometric means at different locations were statistically significant. For private homes, the weighted geometric mean soil lead concentration at the remote location was significantly lower than that at either the entrance or the drip line locations. The differences between the entrance and drip line weighted geometric means are not statistically significant. The weighted geometric mean soil lead concentrations at the drip line, entryway, and remote locations in and around public housing units were also not significantly different. For both private housing and public housing, soil lead concentrations at the three locations were all highly correlated, as shown in Tables 5 and 6 respectively.

Under Section 403 of Title X, EPA has established health-based interim standards for soil lead concentrations and action recommendations for each standard. The agency recommends that "interim controls to change use patterns and establish barriers" should be implemented for areas that are expected to be used by children where soil lead concentrations are between 400 and 5,000 ppm. Within this range, the degree of activity should be "commensurate with the expected risk posed by the bare soil considering both the severity of [lead] exposure...and the likelihood of the children's exposure." For areas where contact by children is less likely or less frequent, the "interim controls" should be implemented when soil lead concentrations are between 2,000 and 5,000 ppm. Moreover, the agency recommends the "abatement of soil" with lead concentrations above 5,000 ppm regardless of the likelihood of children's exposure.

---

[7] Of the 24 soil lead measurements below the instrument detection limit, 4 were entryway soil samples, 8 were drip line samples, and 16 were remote location samples.

Table 3. Detailed distribution of private housing units missing soil lead concentration measurements by age, region, and urbanization

	Missing one, two, or three soil measurements	Missing all three soil lead measurements	Missing no soil lead measurements	Total number of homes in National Survey
Total number of homes	44	18	240	284
Building age				
pre-1940	24	12	53	77
1940 to 1949	6	2	24	30
1950 to 1959	7	2	50	57
1960 to 1979	7	2	113	120
Census region				
Northeast	23	15	30	53
Midwest	8	1	61	69
South	10	2	106	116
West	3	0	43	46
Degree of urbanization				
Urban area in a large metropolitan city	33	14	60	93
Suburban area in a large metropolitan city	7	4	59	66
Urban area in a small metropolitan city	3	0	41	44
Suburban area in a small metropolitan area	1	0	23	24
Nonmetropolitan	0	0	57	57

Table 4. Detailed distribution of public housing units missing soil lead concentration measurements by age and region

	Missing one, two, or three soil measurements	Missing all three soil lead measurements	Missing no soil lead measurements	Total number of homes in National Survey
Total number of homes	72	68	25	97
Building age				
pre-1950	24	22	6	30
1950-1959	20	20	4	24
1960-1979	28	26	15	43
Census region				
Northeast	42	42	1	43
Midwest	5	4	6	11
South	23	21	9	32
West	2	1	9	11

Table 5 Correlations between log-transformed soil lead measurements in private housing from different soil locations around the same dwelling unit

	Soil lead measurements		
	Exterior entrance	Drip line	Remote location
Soil lead entrance	 260	0.7148 0.0001 246	0.6090 0.0001 247
Soil lead drip line	0.7148 0.0001 246	 249	0.6780 0.0001 243
Soil lead remote	0.6090 0.0001 247	0.6780 0.0001 243	 253

Note: For each cell in Table 5, 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 6 Correlations between log-transformed soil lead measurements in public housing from different soil locations around the same dwelling unit

	Soil lead measurements		
	Exterior entrance	Drip line	Remote location
Soil lead entrance		0.7430	0.4313
		0.0001	0.0278
	26	25	26
Soil lead drip line	0.7430		0.7150
	0.0001		0.0001
	25	28	28
Soil lead remote	0.4313	0.7150	
	0.0278	0.0001	
	26	28	29

Note: For each cell in Table 6, 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.

Using the data from the National Survey, an estimated 23 percent, or 18 million, of the private homes in the United States built before 1980 exceed the 400 ppm "further evaluation" guideline; an estimated 6 percent, or almost 5 million, of the private homes in the United States built before 1980 exceed the 2,000 ppm "interim control" guideline; and an estimated 3 percent, or approximately 2.5 million, of the private homes in the United States built before 1980 exceed the 5,000 ppm abatement guideline. Table 7 tabulates the weighted number and percentages of private homes with one or more soil lead concentrations above various levels that might be used as guidelines by EPA. Due to the considerable amount of missing soil samples at public housing units, no national distribution of soil lead prevalence levels is presented for public housing units.

Tables 8 and 9 show estimates of the weighted geometric mean soil lead concentrations for the entryway, drip line, and remote location soil samples by building age, region, and degree of urbanization, for private homes and public housing units. The estimates of the geometric means for public housing presented in Table 9 are not precise due to the small sample sizes ($n<10$) in most of the building age and Census region categories. As a result, the apparent relationships displayed in Table 9 within building age and Census region categories should be interpreted with caution.

3.3 Housing Unit Characteristics

The housing unit characteristics of interest in this study included the building age of the housing unit, the Census region, and degree of urbanization. The construction date and state and county locations of each housing unit were collected by the National Survey and used to classify housing units according to these categories. Using the construction date from the National Survey, each housing unit was classified as being built in one of four time periods for private housing units -- between 1960 and 1979, between 1950 and 1959, between 1940 and 1949, or before 1940 -- and one of three time periods for public housing units -- between 1960 and 1979, between 1950 and 1959, and before 1950. The state in which the housing unit was located was used to classify the housing unit into one of four Census regions: the Northeast, Midwest, South, and West. The regions and the states in each region are shown below:

Census Region	States
Northeast	Maine, New Hampshire, Vermont, Rhode Island, Connecticut, New York, Pennsylvania, New Jersey
Midwest	Ohio, Indiana, Illinois, Michigan, Wisconsin, Minnesota, Iowa, Missouri, Kansas, Nebraska, North Dakota, South Dakota
South	Delaware, Maryland, the District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Mississippi, Alabama, Tennessee, Kentucky, Arkansas, Louisiana, Oklahoma, Texas
West	Montana, Wyoming, Colorado, New Mexico, Arizona, Utah, Idaho, Washington, Oregon, Nevada, California, Hawaii, Alaska

Table 7. Estimated percent and number of U.S. homes built before 1980 exceeding various soil lead concentrations

Soil Lead Concentration (ppm)*	Estimated percent of U.S. homes built before 1980 exceeding the concentration (and 95 percent confidence interval**)	Estimated number (000) of U.S. homes built before 1980 exceeding the concentration (and 95 percent confidence interval**)
400	23.4% (14.7%, 34.4%)	18,090 (11,363, 26,582)
500	20.3% (12.6%, 30.3%)	15,695 (9,746, 23,399)
1,000	11.3% (6.9%, 17.4%)	8,724 (5,329, 13,435)
2,000	7.7% (4.7%, 11.9%)	5,943 (3,661, 9,175)
2,500	6.2% (3.9%, 9.6%)	4,802 (2,984, 7,387)
3,000	3.4% (2.2%, 5.2%)	2,652 (1,706, 3,991)
4,000	3.4% (2.2%, 5.2%)	2,652 (1,706, 3,991)
5,000	3.1% (2.0%, 4.7%)	2,424 (1,569, 3,632)
Total Homes	100%	77,179
Note: Sample Size = 266 homes with data		

* The soil lead concentration is the maximum concentration among the drip line, entrance, and remote location samples for each household with soil lead data.

** The methodology used to calculate the confidence intervals is presented in Section 4.3.

Table 8. Weighted geometric means for soil lead concentrations (ppm) by soil location and housing unit characteristic for private housing units

	Soil Location		
	Drip Line	Entryway	Remote Location
Building age			
pre-1940	480	393	183
1940 to 1949	151	135	67
1950 to 1959	70	74	44
1960 to 1979	27	38	23
Census region			
Northeast	198	161	102
Midwest	109	110	48
South	51	63	38
West	35	58	33
Degree of urbanization			
Urban area in a large metropolitan city	69	88	58
Suburban are in a large metropolitan city	71	78	44
Urban area in a small metropolitan city	130	118	53
Suburban area in a small metropolitan area	64	72	38
Nonmetropolitan	60	77	39
Number of measurements	249	260	253

Table 9. Weighted geometric means for soil lead concentrations (ppm) by soil location and housing unit characteristic for public housing units

	Drip line	Entryway	Remote location	Number of measurements*
Building age				
pre-1950	115	171	131	8
1950-1959	183	184	44	4
1960-1979	31	30	32	6
Census region				
Northeast	45	230	9	1
Midwest	41	34	49	7
South	41	52	30	10
West	97	75	80	10
Number of measurements	28	26	29	

* The number of measurements represents the average number across all soil locations of soil lead-level readings used to estimate the geometric mean.

The housing unit's county and related county statistics were used to designate the unit as belonging to one of five urbanization categories: urban area in a large metropolitan city, suburban area in a large metropolitan city, urban area in a small metropolitan city, suburban area in a small metropolitan city, or nonmetropolitan area. These categories were defined based on i) the size of the Primary Metropolitan Statistical Area (PMSA) or Metropolitan Statistical Area (MSA) in which the county was located and ii) whether or not the county is in the central city of the PMSA or MSA.[8] No such designations were made for public housing units.

Degree of Urbanization	Definition
Urban area in a large metropolitan city	Area located in a central city of a PMSA/MSA with a population of over 1 million.
Suburban area in a large metropolitan city	Area located in a PMSA/MSA with a population of over 1 million, but not located in a central city.
Urban area in a small metropolitan city	Area located in a central city of a PMSA/MSA with a population of less than 1 million.
Suburban area in a small metropolitan city	Area located in a PMSA/MSA with a population of less than 1 million.
Rural/nonmetropolitan area	Area not located in a PMSA/MSA.

Other data derived from the National Survey and included in the analyses were the XRF and lead paint hazard variables. The rationale for including these variables in the model was as follows: 1) to examine the relationship between soil lead and the presence (defined using the XRF variable) and condition (defined using the lead paint hazard variables) of interior and exterior lead-based paint and 2) to control for these factors when assessing the effects of the housing unit characteristics.

The wet and dry room (interior) and exterior XRF variables are the natural logarithms of the average of the XRF readings on all components weighted by the painted surface area of the components in the sampled room. A household average XRF variable was calculated as the arithmetic mean of the wet room, dry room, and exterior XRF variables. The wet and dry room (interior) and exterior lead paint hazard variables are the natural logarithms of the average of the XRF readings on all components weighted by the damaged paint surface area of the components in the sampled room. A household average lead paint hazard variable was calculated as the arithmetic mean of the wet room, dry room, and exterior lead paint hazard variables.

In an attempt to capture the effects of local traffic volume, the National Survey was supplemented with data on traffic in the neighborhoods of the privately owned housing units in the sample. The traffic volume, in vehicle miles per day, was calculated for each housing unit in the following manner: the length of each road within an eighth of a mile of the housing unit was multiplied by the average number of motor vehicles that passed along that road in a 24-hour period, and these products were summed across all roads in the eighth of a mile radius of the dwelling unit.

8 The largest city in each PMSA or MSA is designated a "central city." There may be additional central cities if specified requirements are met. A more complete definition of "central city" can be obtained from the U.S. Office of Management and Budget.

The relationship between a household's traffic volume and its soil lead-levels is expected to be nonlinear. Consequently, the traffic volume data were transformed by centering the natural logarithm of the average daily traffic count at zero to reduce the correlation between the linear and quadratic traffic terms in the soil lead models discussed in Chapter 4. A more complete description of the traffic volume data can be found in *Data Analysis of Lead in Soil and Dust.*[9] Again, no such data were collected for public housing units.

3.4 Preliminary Analyses of Soil Lead Data and Housing Unit Characteristics

Simple correlations (defined by the product moment correlation coefficient r), which can be used to identify potential relationships between housing unit characteristics and soil lead concentrations and are useful tool in the modeling process, are presented in Tables 10 and 11 for private and public and housing, respectively. The results from the correlation tables are descriptive of relationships in the data, but these relationships may not apply to private or public housing in general.[10] The variables are separated divided three categories: soil lead concentrations, housing characteristics, and lead-based paint hazards. The soil lead concentrations are the natural logarithms of the household soil lead levels analyzed throughout the report, the housing characteristics include the number of family units in the development (for public housing), the vehicle miles per day (for private housing), and the decade in which the development was built (for both public and private housing).[11] The lead-based paint hazards include the average household lead hazard and average household XRF variables.[12]

Private Housing

The building characteristic having the strongest relationship with household soil levels is the age of the building (r=0.60,0.60, and 0.55 for drip line, entryway, and remote locations, respectively). The average daily traffic flow, average household lead hazard, and average household XRF reading (which approximate the amount of lead due to traffic, the condition of lead-based paint in the building, and the presence of lead-based paint in the building, respectively) were significantly correlated with the household soil lead levels, although with a smaller correlation than with building age. Additional correlations of interest were the age of the building and the average household lead hazard (r=0.28), the age of the building and the average household XRF reading (r=0.19), and the average household lead hazard and average household XRF reading (r=0.37).

Public Housing

Correlations in the public housing data display results similar to those from the private housing correlation analyses. The building characteristic having the overall strongest relationship with household soil lead levels is the age of the building (r=0.62, 0.53, and 0.28 for drip line, entryway, and remote locations, respectively). The number of family units was significantly correlated with entryway soil lead levels (r=0.53) and slightly correlated with drip line and remote location lead levels (r=0.37 and 0.29 for drip line and remote locations, respectively). The average household paint lead hazard was significantly

---

9 *Data Analysis of Lead in Soil and Dust,* September, 1993. EPA Report number 747-R-93-011.

10 A discussion of the suitability of both the private and public housing data is presented in section 3.5.

11 The data are coded as follows: 2 for homes built between 1970 and 1979, 3 for homes built between 1960 and 1969, 4 for homes built between 1950 and 1959, 5 for homes built between 1940 and 1949, 6 for homes built between 1920 and 1939, and 7 for homes built before 1920.

12 A description of these two variables can be found in section 3.3.

Table 10. Correlations between soil lead concentrations and housing unit characteristics for private housing units

	Soil Lead Concentrations			Building Characteristics		Lead-based paint hazards	
	Drip line	Entryway	Remote location	Average daily traffic flow	Age of building	Average household lead hazard	Average household XRF reading
Drip line				**0.23754** **0.0002** **249**	**0.59942** **0.0001** **249**	**0.30009** **0.0001** **245**	**0.35073** **0.0001** **249**
Entryway		See Table 5 for correlations between soil lead concentrations in public housing units		**0.20262** **0.0010** **260**	**0.59511** **0.0001** **260**	**0.29937** **0.0001** **255**	**0.32922** **0.0001** **260**
Remote location				**0.28047** **0.0001** **253**	**0.54941** **0.0001** **253**	**0.29756** **0.0001** **249**	**0.32499** **0.0001** **253**
Average daily traffic flow	**0.23754** **0.0002** **249**	**0.20262** **0.0010** **260**	**0.28047** **0.0001** **253**		**0.19335** **0.0011** **284**	*** 276	*** 276
Age of building	**0.59942** **0.0001** **249**	**0.59511** **0.0001** **260**	**0.54941** **0.0001** **253**	**0.19335** **0.0011** **284**		**0.27500** **0.0001** **276**	**0.19335** **0.0001** **284**
Average household lead hazard	**0.30009** **0.0001** **245**	**0.29937** **0.0001** **255**	**0.29756** **0.0001** **249**	*** 276	**0.27500** **0.0001** **276**		**0.37416** **0.0001** **276**
Average household XRF reading	**0.35073** **0.0001** **249**	**0.32922** **0.0001** **260**	**0.32499** **0.0001** **253**	*** 276	0.19335 0.0001 284	**0.37416** **0.0001** **276**	

Note: **In each cell of Table 10 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.**

Cells in boldface are significant at the 0.05 level.

***** – the correlation is between -0.10 and 0.10 and the p-value is greater than 0.1.**

Table 11. Correlations between soil lead concentrations and housing unit characteristics for public housing units

	Soil Lead Concentrations			Building Characteristics		Lead-based paint hazards	
	Drip line	Entryway	Remote location	Family units in the building	Age of the building	Average household lead hazard	Average household XRF reading
Drip line				0.36990 0.0527 28	**0.62071** **0.0143** **28**	0.34533 0.0719 28	*** 28
Entryway	See Table 6 for correlations between soil lead concentrations in public housing units			**0.53099** **0.0053** **26**	**0.52885** **0.0055** **26**	**0.49764** **0.0097** **26**	*** 26
Remote location				0.29463 0.1208 29	0.27882 0.1430 29	0.26167 0.1703 29	*** 29
Family units in the building	0.36990 0.0527 28	**0.53099** **0.0053** **26**	0.29463 0.1208 29		0.17447 0.0874 97	*** 97	*** 97
Age of the building	**0.62071** **0.0143** **28**	**0.52885** **0.0055** **26**	0.27882 0.1430 29	0.17447 0.0874 97		*** 97	0.15895 0.1199 97
Average household lead hazard	0.34533 0.0719 28	**0.49764** **0.0097** **26**	0.26167 0.1703 29	*** 97	*** 97		0.18390 0.0714 97
Average household XRF reading	*** 28	*** 26	*** 29	*** 97	0.15895 0.1199 97	0.18390 0.0714 97	

Note: In each cell of Table 11 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.

Cells in **boldface** are significant at the 0.05 level.

*** -- the correlation is between -0.10 and 0.10 and the p-value is greater than 0.2.

correlated with entryway soil lead levels ($r=0.50$) and slightly correlated with drip line and remote locations lead levels ($r=0.35$ and 0.25 for drip line and remote locations respectively). The estimated correlations between average household XRF and soil lead readings, however, were not significantly different from zero.

3.5 Suitability of Soil Lead Data

One important measure of the usefulness of the data is how the distributions of the housing characteristics in the National Survey compare to national distributions. National distributions were obtained from the American Housing Survey for 1987, performed by the Bureau of the Census and HUD for private housing units, and from HUD for public housing units.[13] The distributions of building age and Census region from the National Survey were compared to their respective national distributions. Chi-square tests were used to determine how the distributions in the National Survey compared to those from the American Housing Survey for private homes and the data provided by HUD for public housing units. Variance inflation factors of 1.45 for private housing and 1.13 for public housing units were used to deflate the observed chi-square values to adjust for the survey design effect.[14] Results from the chi-square tests are presented in Table 12 for private homes and Table in 13 for public housing units.

Private Housing Data

For private housing units, the distribution of households in the National Survey was not significantly different from the distribution of households in the American Housing Survey with respect to building age. However, differences with respect to the Census region were marginally significant ($p=0.07$) in that more dwelling units located in the South were sampled in the National Survey than expected based on the American Housing Survey. Because the distributions of households in the National Survey were not significantly different from those found in American Housing Survey and a large amount of data was available (over 250 observations for each model), there are no apparent reasons why inferences cannot be drawn from analyses for private homes.

Public Housing Data

For public housing units, differences between the distribution of sampled public housing units and the distribution of all public housing units, provided by HUD, are significant ($p=0.04$) based on both Census region and building age. Moreover, problems with the lack of soil lead measurements make analyses of the data difficult to interpret. As noted earlier, soil lead samples were taken at only 30 percent (29 of 97) of the sampled public housing units. Given both the distributional inequality and the relatively small number of public housing units where soil samples were taken ($n=29$), all conclusions about public housing units and results from analyses of the public housing data should be viewed with caution.

---

[13] The data used to represent the national distributions of building age and region can be found in the reports of the National Survey, primarily Tables 3-6 and 3-7 of *the EPA Report on the National Survey of Lead-Based Paint in Housing – Appendix II: Analysis.*

[14] The variance inflation factors (VIFs) were estimated in the original analysis of the National Survey data.

Table 12. Chi-square results for building age and Census region variables for private housing units

Building Age	pre-1940	1940 to 1949	1950 to 1959	1960 to 1979
Housing Units Observed from National Survey	77	30	57	120
Estimated from American Housing Survey (1987) (thousands)	21,215	7,945	13,056	36,965
Expected frequencies*	76.1	28.5	46.8	132.6
Individual chi-square values*	0.010	0.079	2.209	1.194

*The chi-square statistic was calculated assuming fixed total of 284 homes with data on building age (4 cells and 3 degrees of freedom).

Total chi-square statistic 2.41
P-value with 3 degrees of freedom 0.49

Census Region	Northeast	Midwest	South	West
Housing Units Observed from National Survey	53	69	116	46
Observed from American Housing Survey (1987) (thousands)	17,618	20,344	25,589	15,628
Expected frequencies**	63.2	73.0	91.8	56.1
Individual chi-square values**	1.644	0.216	6.390	1.804

**The chi-square statistic was calculated assuming a fixed total of 283 homes with data on region (4 cells and 3 degrees of freedom).

Total chi-square statistic 6.93
P-value with 3 degrees of freedom 0.07

Note: The chi-square statistics represent the sum of the individual chi-square statistics weighted by the design effect.

Table 13. Chi-square results for building age and Census region variables for public housing units

Building Age	pre-1950	1950-1959	1960-1979
Housing Units Observed from National Survey	30	24	43
From HUD's national database (thousands)	162	247	388
Expected frequencies*	19.7	30.0	47.3
Individual chi-square values*	5.433	1.213	0.391

*The chi-square statistic was calculated assuming fixed total of homes with data on building age (3 cells and 2 degrees of freedom).

Total chi-square statistic	6.23
P-value with 2 degrees of freedom	0.04

Census Region	Northeast	Midwest	South	West
Housing Units Observed from National Survey	43	11	32	11
From HUD's national database (thousands)	272	152	361	90
Expected frequencies**	30.1	16.8	40.0	10.0
Individual chi-square values**	5.483	2.009	1.618	0.101

**The chi-square statistic was calculated assuming a fixed total of homes with data on region (4 cells and 3 degrees of freedom).

Total chi-square statistic	8.15
P-value with 3 degrees of freedom	0.04

Note: The chi-square statistics represent the sum of the individual chi-square statistics weighted by the design effect.

3.6 Implications of Missing Soil Lead Data

The National Survey protocols specified sampling of soil on the selected property with a soil coring device.[15] Soil samples were not to be collected on neighboring properties if samples could not be collected on the property selected. A percentage of both private housing and public housing buildings (6 and 70 percent respectively) were surrounded by pavement preventing any soil core samples. Two questions arise as a result of the missing soil samples: i) are the soil samples taken representative of the soil samples of interest and ii) how do the missing soil samples affect the results.

Different uses of the data may have required alternative sampling protocols. Some alternative sampling protocols include:

1) Sampling soil in the neighborhood of the housing development, even if only on neighboring properties,

2) Sampling soil as a form of exterior dust in which the dust might be collected using a vacuum or scrape sample from dwelling units with no soil areas, and

3) Sampling the vegetation and/or other soil coverings, as well as the soil to examine the entire lead hazard.[16]

To the extent that the soil samples collected in the National Survey are similar to or representative of the soil samples of interest, the results presented in later sections might be viewed as applicable to public and private housing nationally. According to the 403 Interim Final Rule, soil samples should be taken on bare soil in the area of concern. As a result, the soil samples collected in the National Survey, core soil samples taken on the property, can be viewed as representative of samples called for in the 403 Interim Final Rule.

If soil lead concentrations are higher near older homes, homes in the Northeast region, and homes in large metropolitan urban areas -- the housing unit characteristics associated with the bulk of the missing data -- the estimated impacts of building age, Census region, and degree of urbanization on soil lead concentrations and the estimated number of homes exceeding the various soil lead concentrations would be lower than the *true* impacts and the *true* number of homes, respectively. Since only six percent of the privately-owned homes had no soil areas for soil core sampling, the impact of the missing soil lead data is not expected to be significant and the descriptive statistics in the tables and the results from the analyses can be viewed as applicable to private housing nationally. The results from public housing data, however, should only be viewed as descriptive of those samples collected because i) the sample of public housing is not representative of public housing developments nationally and ii) the impact on the prevalence and distributions of soil lead levels as a result of missing almost 70 percent of the soil lead data is expected to be significant.

---

[15] It should be noted that at housing units where no soil samples were taken scrape sampling might have been possible. Such sampling methods, however, would produce questions concerning the measurement comparability between core and scrape samples. These questions, in turn, would make it difficult to compare the core and scrape sample lead concentration measurements.

[16] If soil has high concentrations of lead from external sources, such as lead in gasoline and lead in exterior or interior paint, it is likely that the vegetation and/or other soil coverings would have high concentrations of lead as well.

4. Statistical Approach

This chapter discusses the modeling and testing procedures used to show the relationship between housing unit characteristics and soil lead concentrations and explains how the confidence intervals for classification percentages were estimated. Many researchers believe that soil lead comes mainly from paint lead and automobile emissions. A review of the evidence in support of this hypothesis can be found in the *Comprehensive and Workable Plan for the Abatement of Lead-Based Paint in Privately-Owned Housing*.[17] Similarly, interior dust levels are believed to be related to soil lead levels.

4.1 Private and Public Housing Model

The purpose of the private and public housing model is to produce estimates of the relative strengths of the associations between the natural logarithm of the soil lead concentrations (response variables) and the housing unit characteristics, XRF measurements, and paint lead hazards (explanatory variables) to determine which of the explanatory variables are good predictors of soil lead. It is to be noted, though, that a strong statistical association between the explanatory (housing unit characteristics and paint lead variables) and response (natural logarithm of the soil lead concentrations) variables does not by itself establish a causal relationship among them. The two variables may have a strong statistical relationship but not a causal relationship. These variables may be caused by a third, unidentified variable, or the relationship may be a statistical artifact.

Assume the following relationship between soil lead levels and housing unit characteristics and other factors affecting soil lead:

(1) $$Y=\alpha+\beta_1X_1+\beta_2X_2+...+\beta_kX_k+\varepsilon$$

In this model, Y represents the response variable, $X_1, X_2, \ldots, X_k$ represent the housing unit characteristics and other factors affecting soil lead, α is the intercept, the parameters $\beta_1, \beta_2, \ldots, \beta_k$ are the coefficients of $X_1, X_2, \ldots, X_k$ respectively, and ε is the measurement error. Having knowledge of the parameters allows the determination of which characteristics or factors play an important role in determining or predicting soil lead concentrations. By combining the categorical characteristics and factors into T and the assigning the leftover n ($n<k$) continuous characteristics and factors as $V_1, V_2, \ldots, V_n$, the model can be rewritten as an analysis of covariance model:

(2) $$Y=\alpha+\gamma T+\delta_1V_1+\delta_2V_2+...+\delta_nV_n+\varepsilon$$

The parameter γ is a vector containing the parameters of all the categorical variables, and the parameters $\delta_1, \delta_2, \ldots, \delta_n$ correspond to the parameters for the continuous variables in model (1). The method of weighted least squares can be used to obtain estimates of α, γ, and $\delta_1, \delta_2, \ldots, \delta_n$ by fitting the observed values of Y to the observed values of T and $V_1, V_2, \ldots, V_n$:

(3) $$Y=a+cT+d_1V_1+d_2V_2+...+d_nV_n+e$$

[17] *Comprehensive and Workable Plan for the Abatement of Lead-Based Paint in Privately Owned Housing: Report to Congress*. December 7, 1990. U.S. Department of Housing and Urban Development. Washington DC.

The coefficients a, c, and $d_1, d_2, \ldots, d_n$ are the estimates of the model parameters α, γ, and $\delta_1, \delta_2, \ldots, \delta_n$ and are calculated so that the weighted variance of the prediction errors, or residuals, e, is minimized. The weights in the model were the sampling weights. As a result of the sample design, a variance inflation factor is applied to the variance estimates to generate unbiased estimates.

The parameter estimates will be unbiased estimates of the true parameters if all three of the following conditions hold: 1) the natural logarithm of the soil lead, 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 model has the same independent variables and mathematical form as the true relationship. Biased parameter estimates could lead to incorrect conclusions about the relationships between soil lead concentrations and housing unit characteristics.

Although it is likely that some, if not all, of the continuous explanatory variables are measured with error, the lack of knowledge about the true relationship between the explanatory and response variables is the most important concern with respect to these models. Because of this lack of knowledge, it is important to keep all variables in the analysis that might affect the response variable. If key explanatory variables are left out, the estimates of the response variable based on the remaining explanatory variables may be biased. If extra explanatory variables are included in the model, the model estimates for the true explanatory variables will be unbiased, but only in the absence of measurement error in the independent variables. The parameter estimates, though, will not be as precise as if the extraneous variables were not in the model.

In the analysis of covariance model, parameter estimates are generated for all variables. These estimates for continuous variables are unbiased (if all three of the above criteria are met) and have simple interpretations. For these variables, the parameter estimates and 95 percent confidence intervals are reported. The statistical significance of the categorical variables and the least squares means for each level within a categorical variable are reported. The least squares means are estimates of the average response (soil lead concentration) given the particular classification of the categorical variable of interest, while holding all other variables at their averages.

4.2 Modeling and Testing Procedures

All variables that conceptually have a significant impact on household soil lead levels and were available in the National Survey database were used in the initial analyses. These variables included the building or development's age (measured as the date of construction), Census region, and degree of urbanization (for private housing), two-way interactions between the age and Census region, the Census region and degree of urbanization (for private housing), and the age and degree of urbanization (for private housing), a three-way interaction between the age, Census region, and degree of urbanization, the building's average daily traffic flow (for private housing), the number of units in the development (for public housing), and interior and exterior XRF and lead hazard variables,[18] which approximate the presence and condition of lead-based paint, respectively.

Because parameter estimates in models with extraneous variables are imprecise due to inflated variances estimates, the extraneous variables in the soil location models were removed. Methods for removing extraneous variables range from keeping all possibly relevant terms, regardless of their

[18] An aggregated household average XRF and lead hazard variable replaced the interior and exterior XRF and lead hazard variables.

statistical significance, to keeping all significant terms regardless of their relevance. A method which strikes a balance between these two bounds was used. The key variables of interest to the study--building age, Census region, and degree of urbanization (for private housing only)--were always kept in the model regardless of their statistical significance. Then, the most statistically insignificant variables were removed one at a time, unless they were one of the key variables of interest in the study. Variables that were significant in other soil location models were also kept to create more comparable models. As a result, reasonably relevant terms with some degree of statistical significance, and terms significant in any of the other soil location models, were kept in the final soil lead models.

A factor was considered a significant predictor of household soil lead if it was significant at the 5 percent level in the model fit and the overall regression F statistic was significant at the 5 percent level. In all cases, the overall regression F statistic was significant at the 5 percent level. Levels within factors were considered significantly different if the factor was significant at the 5 percent level in the model fit and the difference between levels was significant at the 5 percent level. No other multiple comparison procedure was used to evaluate differences in the factor levels.

For significant factors, differences among levels were discussed without stating statistical significance. Differences that were not significant were occasionally discussed, but only within the context of understanding the results of the model fit.

4.3 Confidence Intervals for Classification Percentages

The confidence intervals for the percentages reported in Table 7 were estimated using a series of equations that accounted for measurement error, misclassification error due to measurement error (the error associated with improper classifications of soil lead), and the expected asymmetry of the confidence intervals. These calculations were performed for each of the concentration bounds presented in Table 7 as described below.

The first step was to compute the misclassification error, σ_e^2, which was obtained in the following manner:

(4) $$\sigma_e^2 = (\Sigma_i p_i * (1-p_i)) / n^2 \qquad i=1,\ldots,n$$

The value p_i is the probability that the observed maximum soil lead level is greater than the specified concentration limit assuming a normal distribution with the mean equal to the observed maximum value and the variance equal to 0.84.[19] Further, n is the number of homes with at least one soil lead level observation. The variance of the proportion, σ_p^2, can be estimated using the misclassification error as:

(5) $$\sigma_p^2 = (1.45 * p * (1-p)) / n + \sigma_e^2,$$

where p is the observed proportion of homes in the survey with soil lead levels greater than the concentration limit and 1.45 is a variance inflation factor used to adjust variance of the proportion.

[19] This value is the square of the estimated standard deviation of one soil lead measurement. The calculations for this estimate are found in *Data Analysis of Lead in Soil and Dust.*

To generate an asymmetric confidence interval, the proportions, p; are transformed into variables, y, which are approximately normally distributed. The transformation is

(6) $$y(p) = \arcsin(\sqrt{p}).$$

A 95 percent confidence interval for the transformed variables is calculated as

(7) $$y(p) \pm 1.96 * \sigma_y ,$$

where σ_y^2 is the variance of the transformed variables and is calculated as

(8) $$\sigma_y = \partial y/\partial p * \sigma_p = (\partial \arcsin(\sqrt{p}))/\partial p * \sigma_p.$$

Asymmetric lower and upper confidence limits for the proportion p are calculated from the lower and upper confidence limits of y using equation (6).

5. Modelling Results

Soil lead concentrations were regressed on housing unit characteristics, including the building's age, Census region, and degree of urbanization, and the presence and condition of lead-based paint. Additional variables also considered to be related to soil lead and used in the analyses included the average daily traffic flow in the neighborhood of the housing unit (for private housing only) and the number of family units in the development (for public housing only). Soil lead concentrations from each location, the drip line, entryway, and remote location, were analyzed separately. The natural logarithms of the soil lead concentrations were used in all analyses as the response variables. In addition to examining the relationship between soil lead levels and housing unit characteristics, this report also examines the relationships between soil lead levels and interior and exterior paint lead levels.[20]

In each of the soil lead models, soil lead levels were regressed on the housing unit's region, building age, degree of urbanization, building age by region interaction, building age by degree of urbanization interaction (for private housing), average paint lead hazard, average XRF, average daily traffic counts (for private housing), and the number of family units in the building (for public housing). In this section, the results from the analysis of covariance models for the drip line, entryway, and remote locations are presented in Tables 14 through 16 for private housing and Tables 17 and 18 for public housing. These results include the significance and least-squares means of the categorical variables, the parameter estimates of the continuous variables, and the model statistics.

There are two important concepts to remember in the discussions of the results. First, the significance levels of the categorical variables show whether or not the levels of a categorical variable have significantly different effects on the soil lead concentration. Second, the least-squares means show how the levels of a categorical variable differ with respect to their effects on the soil lead concentration.

The categorical and continuous variables that are statistically significant at the 5 percent level are shown in boldface in Table 14 for the private housing results and Table 17 for the public housing results. At the bottom of these tables, the model statistics, the number of observations used in the analysis and the R-square, are presented for each soil lead location model. In these analyses, the R-square is viewed as the percent of variation explained by the model, not as a measure of comparison between models. The least-squares means and 95 percent confidence intervals for the categorical variables in each private housing soil lead model are presented in Figures 1 through 4 and Tables 17 and 18. The least-squares means and 95 percent confidence intervals for the categorical variables in each public housing soil lead model are presented in Figures 5 and 6 and Table 17. Simple correlations between housing unit characteristics, paint lead hazards, and soil lead concentrations in public housing units are presented in Table 15.

The variance estimates from the analysis of covariance models, the mean-square error, and variances of the parameter estimates were inflated as a result of using the sampling weights in the analysis. The private housing variance estimates were inflated by a factor of 1.45 and the public housing units were inflated by a factor of 1.13.

---

[20] Additional discussions and conclusions on the relationship between soil lead levels and paint lead levels can be found in *Data Analysis of Lead in Soil and Dust.*

5.1 Private Housing Results

The strongest predictor of soil lead for all soil sample locations was the age of the dwelling unit. Dwelling unit age measures the length of time since the construction of the building and, in most cases, the last major disturbance of soil. Thus, the dwelling unit age measures the length of time that lead deposits -- from dwelling unit and neighboring activity sources -- have accumulated in the soil. In addition, a two-way interaction involving the building age and Census region was significant in one of the soil lead models. This two-way interaction, building age by region, provides a useful tool to quantify the extent to which the factors of interest are not additive. The least squares means, the estimated average of the soil lead measurements from a soil lead model, and 95 percent confidence intervals for the building age, Census region, and degree of urbanization variables are presented in Figures 1, 2, and 3 respectively and in tabular form in Table 15. The least squares means for the interaction of building age by region are presented graphically in Figure 4 and in tabular form in Table 16.

There were other significant predictors of soil lead in each of the soil location models. These included the Census region, the presence of lead-based paint (as measured by the average XRF reading), and the average daily traffic count. Which predictors were significant depended on the location from which the soil samples were obtained. Although the degree of urbanization was not a significant predictor, it was left in the three soil lead location models because it was one of the key variables of the study. The significant predictors in the drip line and entryway soil lead models were nearly identical, but were different from the remote location soil lead model.

Drip Line and Entryway Models

For both the drip line and entryway soil lead models, the Census region factor was statistically significant, although more significant in the drip line soil lead model than the entryway soil lead model. The building age by Census region interaction was not significant in either the drip line or entryway soil lead models. In both models, the housing units in the Northeast region were shown to have significantly higher soil lead concentrations than soil lead concentrations in the South and West regions and have higher soil lead concentrations than the Midwest region after adjusting for the housing unit's age.

Many studies have shown that urban areas have higher soil lead concentrations than suburban and rural areas.[21] In this analysis, it was expected that homes in urban areas would have higher soil lead concentrations than homes in suburban and rural areas. Similarly, homes in large metropolitan areas would have higher soil lead concentrations than homes in small metropolitan areas. In the drip line and entryway soil lead models, the degree of urbanization factor was not significant. As a result, soil lead levels around homes in urban, suburban, and rural areas are not significantly different.

There are a number of possible explanations for this unanticipated result. One explanation might be found in reviewing the distribution of the missing soil lead measurements. Generally, soil lead concentrations are expected to be higher in large, highly urbanized areas. However many such sites have very little, if any, soil. The larger and more urbanized a site, and the more likely the soil is to have high lead concentrations, the more likely it is that the soil has been paved over. As a result, average soil lead

---

21 Examples of such studies include HW Mielke, et al, "Lead concentrations in the inner-city soils as a factor in the child lead problem," American Journal of Public Health, 1983, and ID Shellshear, et al, "Environmental Lead Exposure in Christchurch children: soil lead a potential hazard," New Zealand Medical Journal, 1975.

Table 14 Soil lead model statistics for private housing models

	Soil Location		
	Drip Line	Entryway	Remote Location
Significance of the Categorical Variables			
Building age	.0001	.0001	.0001
Census region	.01	.08	.06
Degree of urbanization	**	**	**
Building age by Census region	**	**	.006
Parameter Estimates and 95 Percent Confidence Intervals for the Continuous Variables			
Average household XRF reading	0.036 (0.014,0.057)	0.042 (0.023,0.062)	0.030 (0.009,0.051)
Traffic	-0.861 (-1.371,-0.351)	-1.242 (-1.703,-0.780)	1.140 (0.660,1.619)
Traffic squared	0.091 (0.045,0.137)	0.112 (0.071,0.154)	-0.081 (-0.506,0.344)
Model Statistics			
R-Square	.611	.542	.513
Number of observations	249	260	253

** - not significant at the 0.10 level

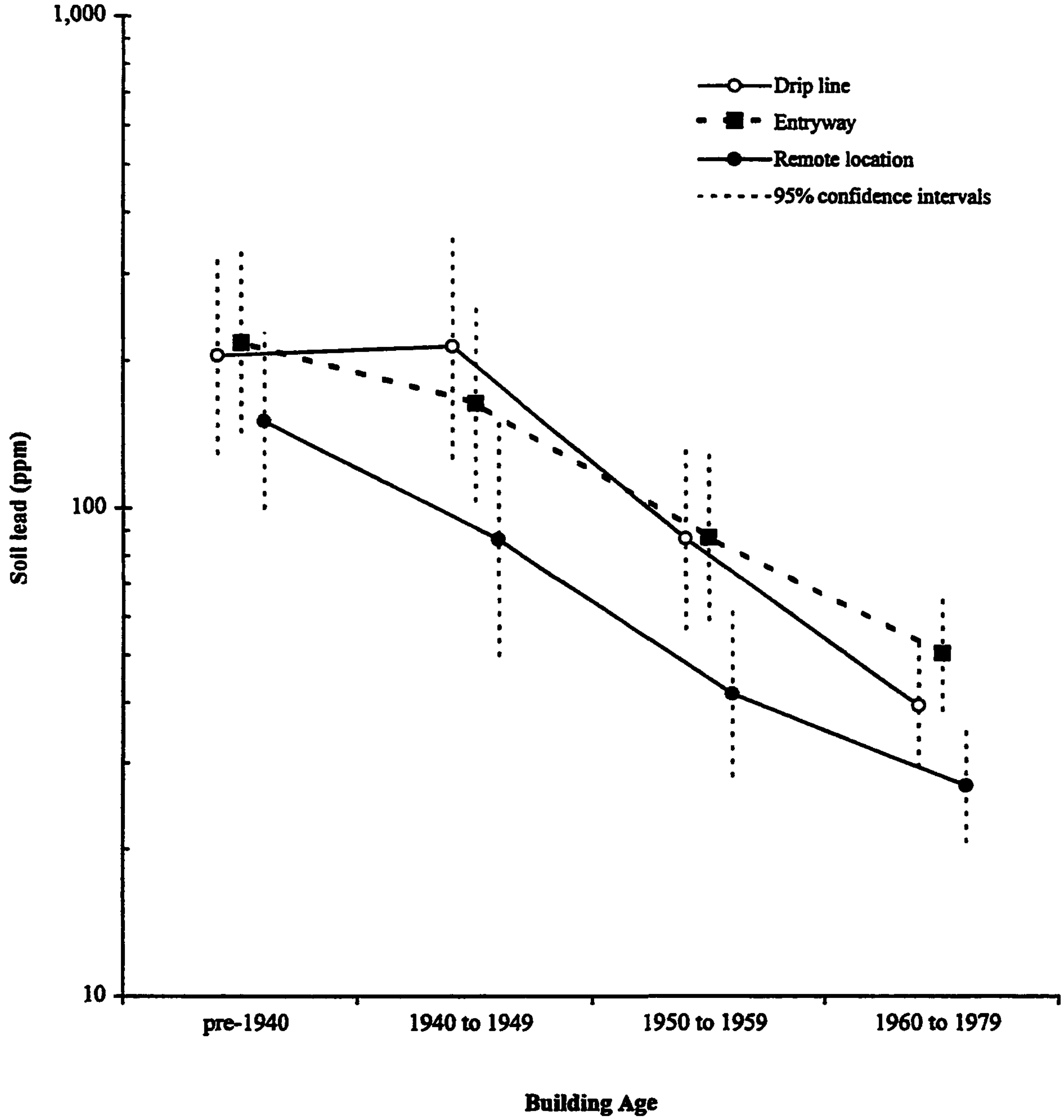

Figure 1. Least squares means and 95 percent confidence intervals for soil lead concentrations in private housing for building age by soil location

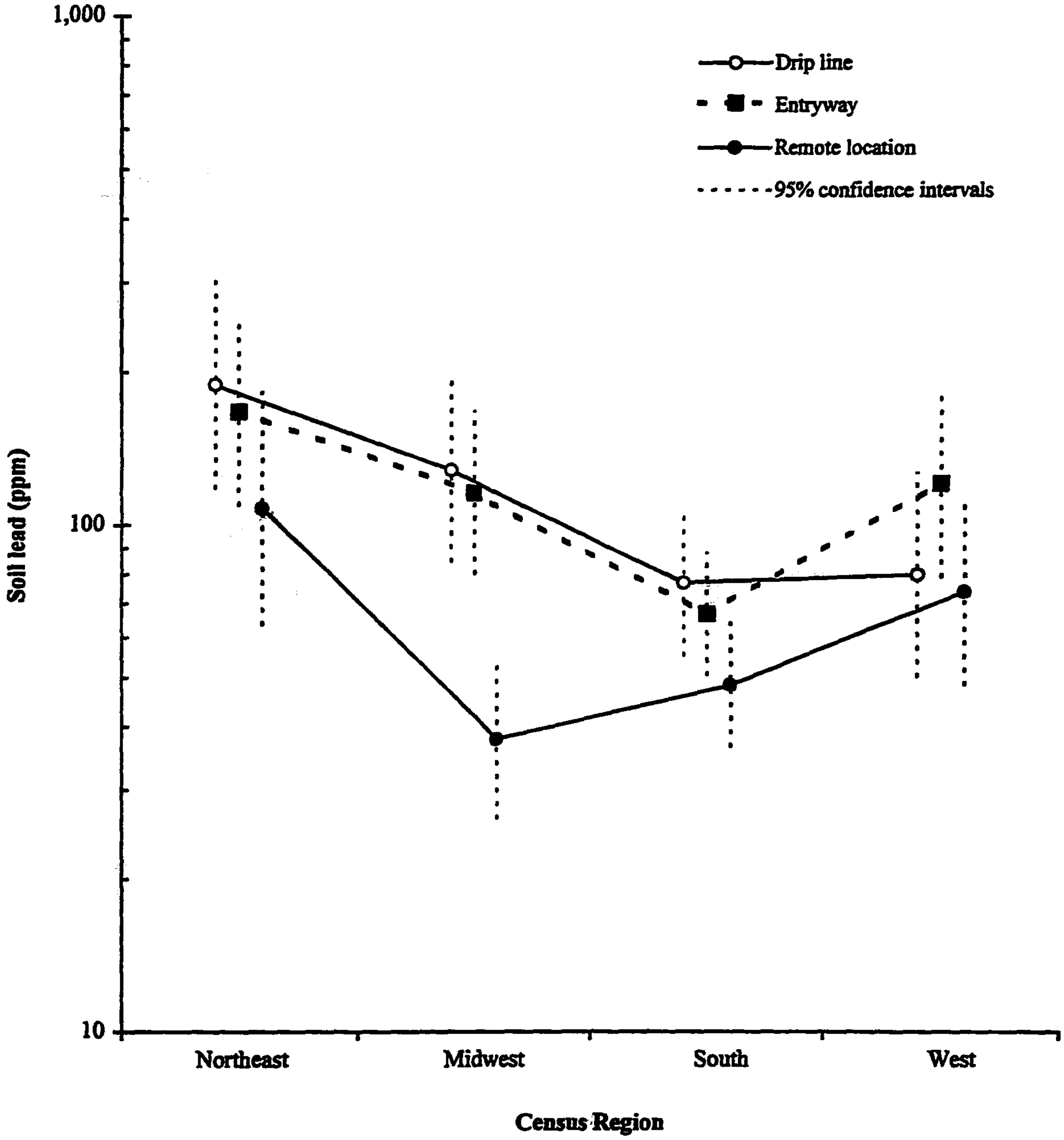

Figure 2. Least squares means and 95 percent confidence intervals for soil lead concentrations in private housing for Census region by soil location

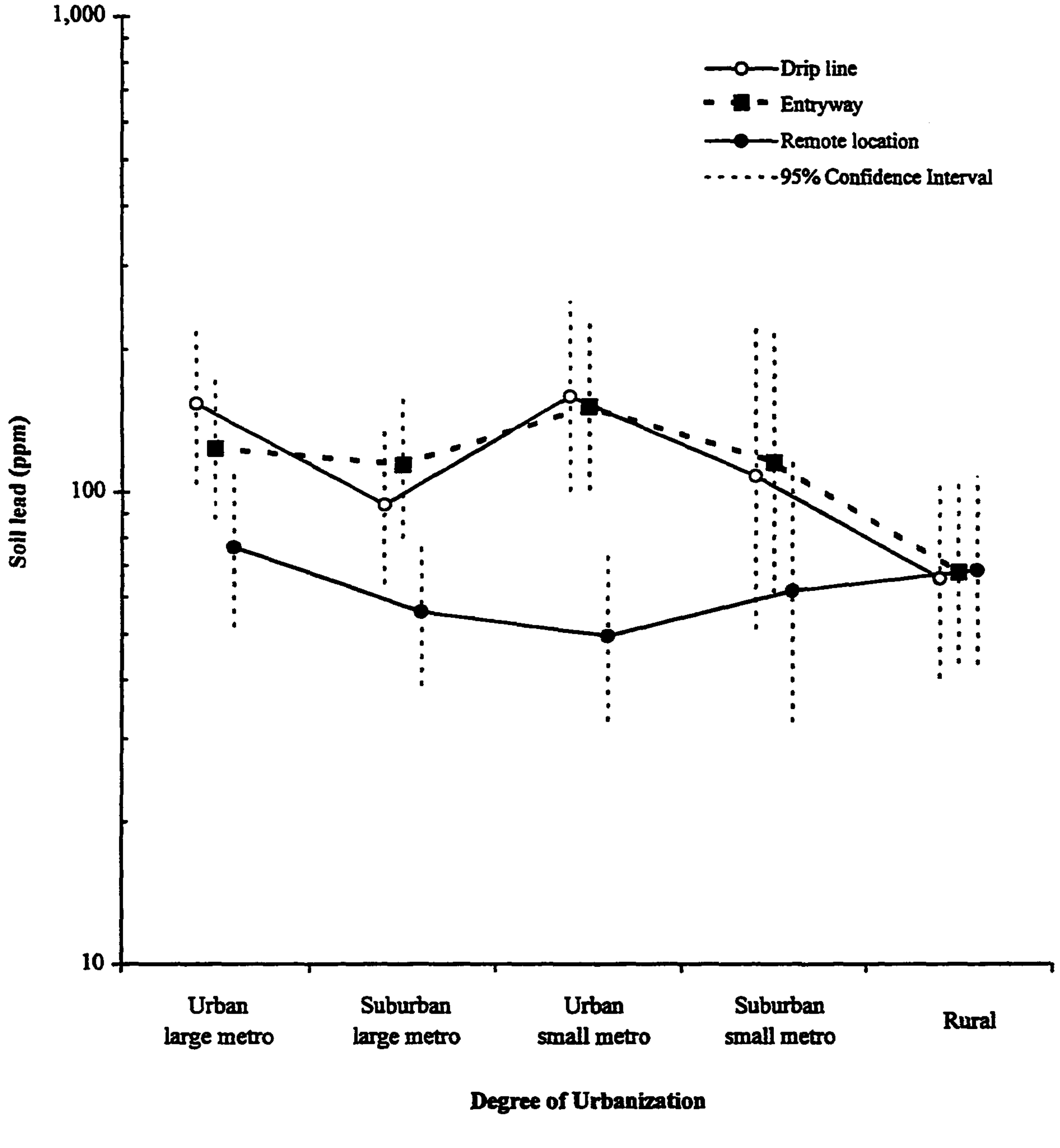

Figure 3. Least squares means and 95 percent confidence intervals for soil lead concentrations in private housing for degree of urbanization by soil location

Table 15. Least-squares means and 95 percent confidence intervals (ppm) for categorical variables in the private housing unit models

	Soil Lead Model		
	Drip Line	Entryway	Remote Location
Building age			
1960-1979	36.7 (28.9, 46.6)	47.0 (37.7, 58.7)	25.6 (20.3, 32.2)
1950-1959	75.1 (52.5, 107.6)	85.2 (61.5, 118.2)	39.4 (27.9, 55.5)
1940-1949	157.1 (95.5, 258.5)	152.7 (97.6, 239.1)	90.1 (55.4, 146.7)
pre-1940	329.5 (234.0, 463.8)	256.5 (189.1, 348.0)	210.1 (151.9, 290.6)
Census region			
Northeast	177.1 (116.0, 270.3)	157.2 (107.4, 230.1)	117.8 (76.5, 181.4)
Midwest	147.5 (102.3, 212.8)	125.1 (90.1, 173.6)	50.6 (36.0, 71.1)
South	84.0 (60.9, 115.9)	77.2 (58.4, 102.0)	50.9 (38.0, 68.3)
West	65.0 (42.5, 99.2)	103.5 (70.1, 153.0)	62.9 (42.0, 94.3)
Degree of urbanization			
Urban area in a large metropolitan area	103.8 (73.9, 145.7)	110.4 (81.9, 148.7)	78.7 (57.0, 108.6)
Suburban area in a large metropolitan area	95.9 (70.3, 130.7)	104.0 (78.5, 137.7)	55.0 (40.8, 74.2)
Urban area in a small metropolitan area	145.7 (96.4, 220.0)	137.7 (95.6, 198.3)	53.6 (36.3, 79.2)
Suburban area in a small metropolitan area	142.3 (87.4, 231.6)	140.6 (91.2, 216.8)	66.8 (42.4, 105.3)
Nonmetropolitan	75.6 (50.3, 113.6)	79.2 (54.6, 114.9)	81.5 (55.1, 120.5)

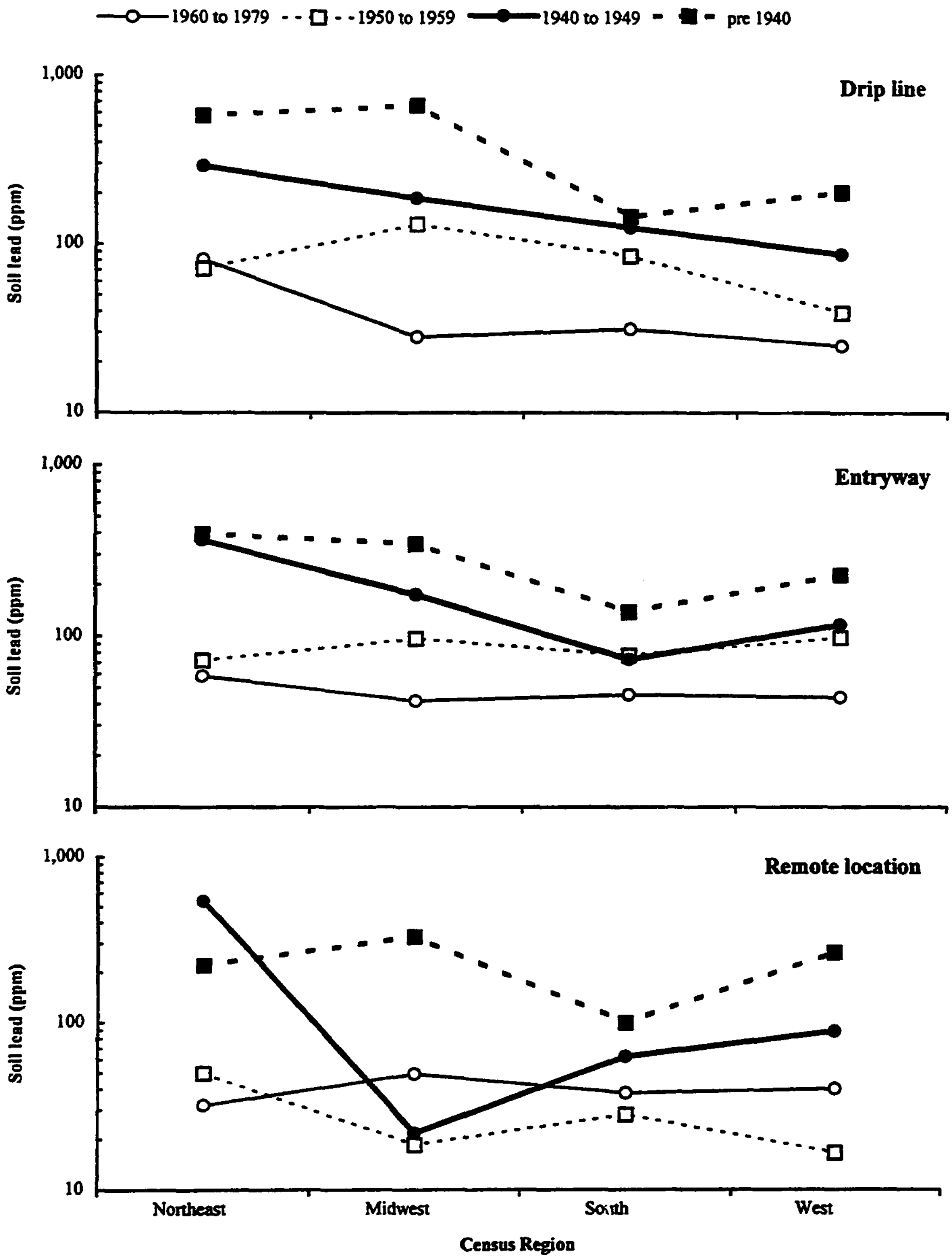

Figure 4. Least squares means for soil lead concentrations in private housing for the building age and Census region interaction by soil location

Table 16. Least-squares means and 95 percent confidence intervals for building-age by Census region interactions in the private housing unit models

Housing Unit Characteristic		Soil Lead Concentrations (ppm)		
Building age	Census region*	Drip line soil lead	Entryway soil lead	Remote soil lead
1960-1979	NE	81.2 (43.4,152.0)	58.6 (33.1,103.7)	49.7 (27.3,90.5)
	MW	28.2 (17.8,44.7)	41.8 (27.6,63.5)	18.5 (11.9,28.7)
	S	31.5 (22.9,43.3)	45.5 (34.2,60.5)	28.2 (20.8,38.1)
	W	25.1 (15.7,40.1)	43.9 (28.1,68.8)	16.6 (10.6,26.0)
1950-1959	NE	71.7 (34.0,151.0)	72.1 (36.6,141.8)	32.1 (15.5,66.4)
	MW	132.1 (64.3,271.1)	96.7 (51.0,183.6)	49.3 (24.8,98.0)
	S	85.5 (47.6,153.5)	77.1 (45.6,130.5)	37.9 (22.0,65.5)
	W	39.4 (18.3,84.7)	98.2 (47.8,201.7)	40.1 (19.3,83.3)
1940-1949	NE	291.2 (90.5,937.2)	365.1 (126.3,1054.9)	543.1 (154.6,1908.0)
	MW	188.5 (72.8,488.4)	174.8 (73.6,415.2)	21.8 (9.3,50.9)
	S	126.3 (53.3,299.3)	73.1 (34.9,153.2)	62.8 (28.9,136.6)
	W	87.9 (31.3,246.7)	116.7 (45.6,298.5)	88.9 (33.1,238.4)
pre-1940	NE	580.0 (322.6,1042.8)	396.7 (239.0,658.3)	222.0 (125.6,392.6)
	MW	674.5 (373.4,1218.6)	345.9 (205.7,581.8)	330.4 (188.8,578.2)
	S	146.7 (78.2,275.5)	138.2 (79.3,240.8)	100.0 (55.8,179.3)
	W	205.3 (89.2,472.7)	228.2 (107.0,486.9)	265.3 (119.7,587.8)

* NE - Northeast
MW - Midwest
S - South
W - West.

concentrations in large metropolitan urban areas (which are missing- at least one soil lead measurement at 33 of 93 sampled units) were found to be lower than those in small metropolitan areas (which are missing at least one soil lead measurement at only 4 of 68 sampled units). A second explanation might be that the correlation between the degree of urbanization and other factors, such as traffic, is reducing the effect of highly urbanized areas. The unanticipated result might also be simply due to the variation in the data associated with the random selection of the homes.

The parameter estimates of the remaining significant predictors of soil lead were relatively consistent across the drip line and entryway models. The parameter estimate for the average XRF reading variable was relatively consistent across both the drip line and entryway soil lead models. In addition, the parameter estimates of both the linear and quadratic terms of the traffic variables were significant and similar in magnitude across both models. Therefore, the relationship between the log transformed traffic variable and log transformed soil lead response variable is nonlinear.

Remote Location Model

For the remote soil location model, as well as in the drip line and entryway models, the housing units in the Northeast region have significantly larger soil lead concentrations at the sampled remote location than do the other regions. In addition, the building age by region interaction was significant. As with the drip line and entryway models, the average household XRF reading variable was a significant predictor of soil lead concentration. The effect of traffic was different, however, in that it was linear in the remote location model.

5.2 Public Housing Results

As discussed in Sections 3.5 and 3.6, problems with the public housing data limit inferences that can be drawn from an analysis of the public housing soil data. The results from the analyses of covariance, presented in Table 17, are descriptive of relationships in the data, but these relationships may not apply to public housing in general.

The results from the analysis of covariance are somewhat different than those from the correlation analysis. The building age was significant in both the analysis of covariance and correlation analysis. However, the average household lead hazard variable and the number of family units were significant in correlation analysis but not significant in the analysis of covariance. The average household XRF reading variable was not significant in either the analysis of covariance or the correlation analysis. These three variables--the number of family units, the average household lead hazard, and the average household XRF reading--do not explain any additional variation in the soil lead concentrations in the presence of the building age and Census region. The analysis of covariance results are presented in Table 17 and least squares means for building age and Census region are presented in Figures 5 and 6, respectively, and in tabular form in Table 18.

When viewing the correlations and the results from the analysis, the reader should be remember that data from only 30 percent of the sampled units were used to estimate the correlations between household soil lead concentrations and development characteristics and lead-based paint hazard variables.

Table 17. Soil lead model statistics for public housing models

	Soil Location		
	Drip Line	Entryway	Remote Location
Significance of the Categorical Variables			
Building age	**.0003**	**.002**	**.009**
Census region	**.04**	**	**.04**
Parameter Estimates and 95 Percent Confidence Intervals for the Continuous Variables			
Average household paint lead hazard	0.232 (-0.240,0.707)	0.417 (-0.124,0.958)	-0.026 (-0.473,0.421)
Average household XRF reading	-0.037 (-0.116,0.042)	-0.051 (-0.137,0.036)	-0.042 (-0.117,0.032)
Model Statistics			
R-Square	.601	.517	.461
Number of observations	27	25	28

** - not significant at the 0.10 level

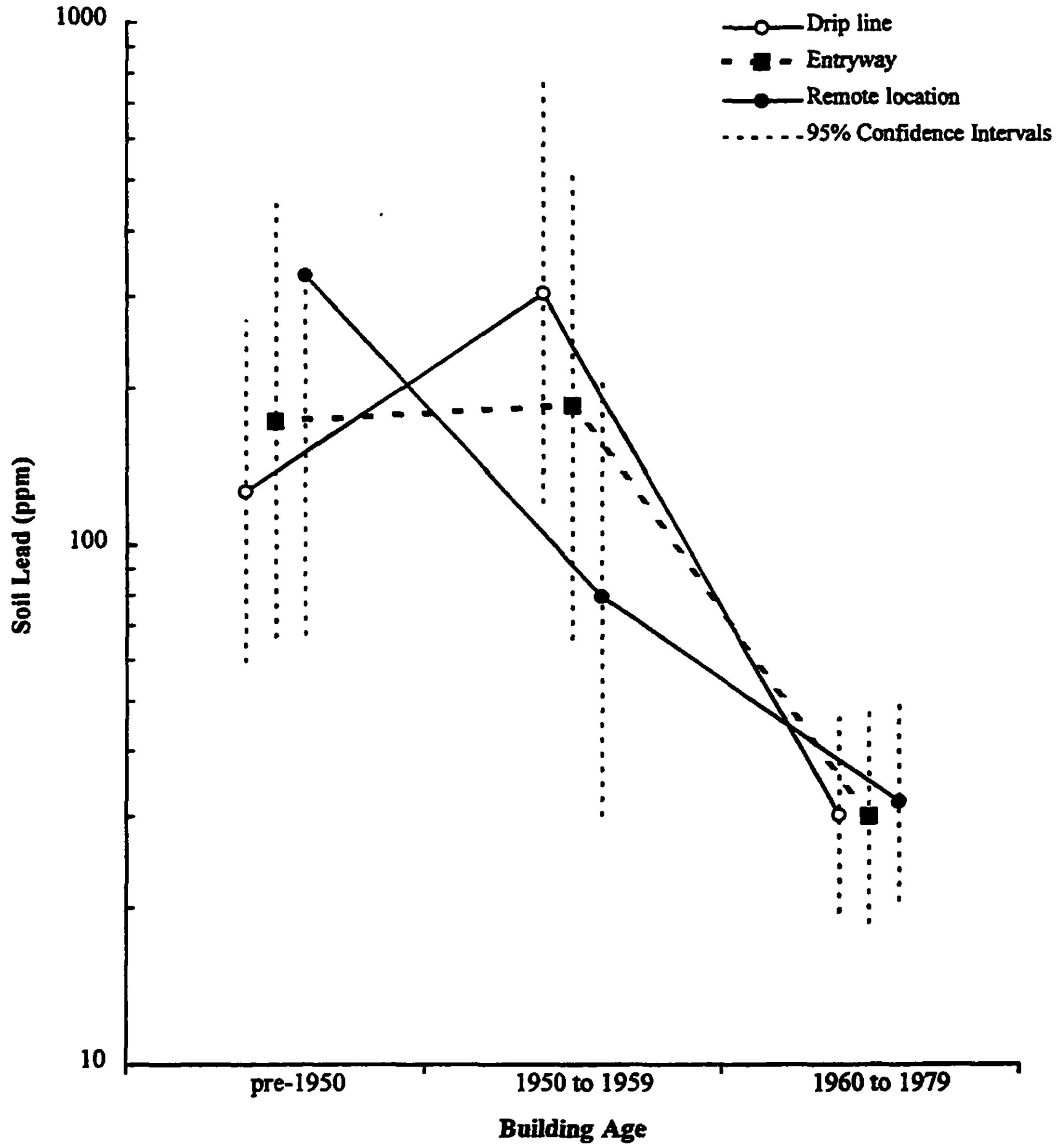

Figure 5. Least squares means and 95 percent confidence intervals for soil lead concentrations in public housing for building age by soil location

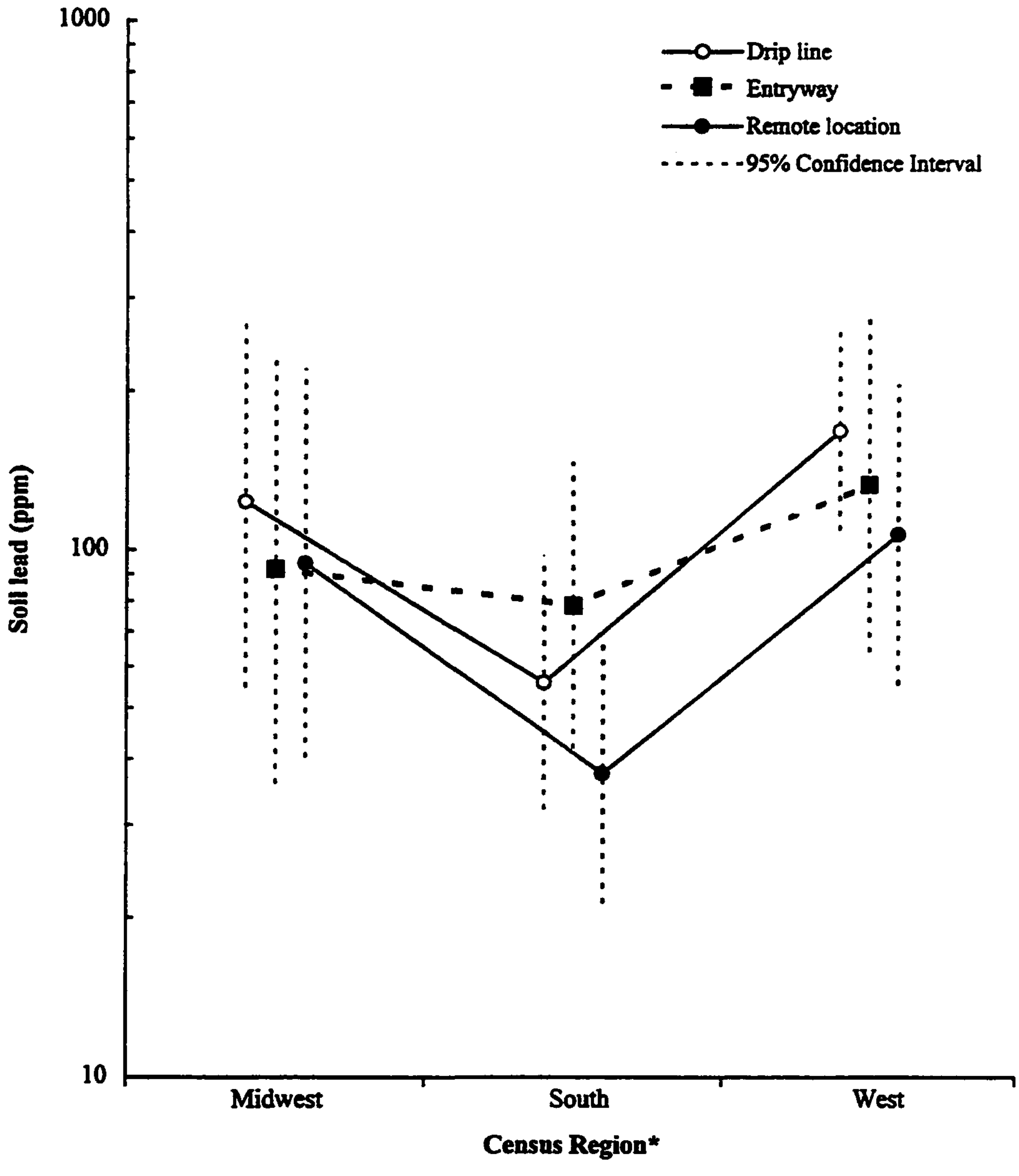

Figure 6. Least squares means and 95 percent confidence intervals for soil lead concentrations in public housing for Census region* by soil location

* No least-squares means were generated form the Northeast region because the one sampled public housing unit with soil lead data was removed from the analysis

Table 18. Least-squares means and 95 percent confidence intervals for categorical variables in the public housing unit models

	Drip line soil lead (ppm)	Entryway soil lead (ppm)	Remote soil lead (ppm)
Building age			
1960-1979	30.2 (19.7,46.4)	30.1 (18.7,48.3)	32.1 (20.7,49.8)
1950-1959	305.5 (121.2,770.2)	186.4 (66.3,524.0)	79.8 (30.2,210.9)
pre-1950	126.8 (59.6,269.7)	173.2 (66.6,450.3)	149.1 (67.4,330.0)
Census region*			
Midwest	123.3 (55.1,276.2)	92.0 (36.1,234.1)	94.2 (40.5,219.1)
South	56.0 (32.3,97.1)	78.7 (42.2,146.8)	37.7 (21.5,66.1)
West	169.3 (110.3,259.9)	134.1 (64.7,278.0)	107.7 (56.0,207.1)

* No least-squares means were generated form the Northeast region because the one sampled public housing unit with soil lead data was removed from the analysis

References

Brown, S.F., Schultz, B., Clickner, R.P., and Weitz, S., August 1992, *Data Analysis of Lead in Soil.* Presented at the American Chemical Societies Annual Meeting.

H.W. Mielke, et al, "Lead concentrations in the inner-city soils as a factor in the child lead problem," *American Journal of Public Health,* 1983

I.D. Shellshear, et al, "Environmental lead exposure in Christchurch children: Soil lead a potential hazard," *New Zealand Medical Journal,* 1975.

Midwest Research Institute, *Analysis of Soil and Dust Samples for Lead (Pb), Final Report,* May 8, 1991. Prepared under contract to the U.S. Environmental Protection Agency. EPA Contract No. 68-02-4252.

U. S. Department of Housing and Urban Development, *Comprehensive and Workable Plan for the Abatement of Lead-Based Paint in Privately Owned Housing: Report to Congress.* December 7, 1990. Washington DC.

U. S. Environmental Protection Agency, *Data Analysis of Lead in Soil and Dust,* September 1993. EPA Report No. 747-R-93-011.

U. S. Environmental Protection Agency, *Report on the National Survey of Lead-Based Paint, Base Report,* June 1995. EPA Report No. 747-R95-003.

U. S. Environmental Protection Agency, *Report on the National Survey of Lead-Based Paint, Appendix I: Design and Methodology,* June 1995. EPA Report No. 747-R95-004.

U. S. Environmental Protection Agency, *Report on the National Survey of Lead-Based Paint, Appendix II: Analysis,* June 1995. EPA Report No. 747-R95-005.

U. S. Environmental Protection Agency, *Guidance on Identification of Lead-Based Paint Hazards,* Federal Register, v 60 (175): September 11, 1995.

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. EPA 747-R-96-003	2.	3. Recipient's Accession No.
4. Title and Subtitle Distribution of Soil Lead in the Nation's Housing Stock			5. Report Date -May, 1996 6.
7. Author(s) Westat, Inc.			8. Performing Organization Report No.
9. Performing Organization Name and Address Westat, Inc. 1650 Research Boulevard Rockville, MD 20850			10. Project/Task/Work Unit No. 11. Contract (C) or Grant (G) No. (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 EPA and HUD, lead measurements were collected on exterior soil, interior house dust, and in interior and exterior paint for each sampled dwelling unit. In addition, the dwelling unit's age, Census region, and degree of urbanization were obtained. This report presents findings from the National Survey on the prevalence and concentrations of lead in soil in private and public housing units in the United States. These findings include national estimates of the number of private housing units with various soil lead concentrations and average soil lead concentrations by building age, Census region, and degree of urbanization. The report also summarizes the statistical associations between soil lead concentrations and building age, degree of urbanization, Census region, and the presence and condition of lead-based paint. An analysis of covariance model was used to identify possible predictors of lead in soil. The age of the dwelling unit was the predominate predictor of soil lead. Other statistically significant predictors of soil lead included the dwelling unit's Census region, the dwelling units' average lead paint levels, and local automobile emissions.

17. Document Analysis

a. Descriptors

Environmental Contaminants

b. Identifiers/Open-Ended Terms

Soil lead, lead-related hazards, National Survey of Lead-Based Paint, Title X, Section 403.

c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report) Unclassified	21. No. of Pages 62
Available to the public from NTIS, Springfield, VA	20. Security Class (This Page) Unclassified	22. Price

(See ANSI-239.18) *See Instructions on Reverse* OPTIONAL FORM 272 (4-77) (Formerly NTIS-35) Department of Commerce