APPENDIX A

Glossary

APPENDIX A

GLOSSARY

[*] **Abatement**: Any set of measures designed to permanently eliminate lead-based paint hazards in accordance with standards established by Federal agencies.

***Accessible or Chewable Surface**: An interior or exterior surface painted with lead-based paint that is accessible for a young child to mouth or chew.

Arithmetic Mean: The sum of a set of measurements divided by the number of measurements.

Background Lead Exposure: Exposure to environmental lead that is not the result of human activity such as lead-based paint or industrial sources.

Baseline: Conditions prior to implementing interventions in response to §403 rules. Baseline risk characterization is performed in this risk analysis using blood-lead concentration data from Phase 2 of NHANES III and by assumptions on the relationship between blood-lead concentration and IQ score decrement.

Biokinetics: Processes affecting the movement of molecules from one internal body compartment to another, including elimination from the body.

Blood-Lead Concentration: Blood-lead concentration measures the mass of lead collected per volume of whole blood collected and is usually expressed in terms of micrograms of lead collected per deciliter of blood collected (µg Pb/dL blood).

Blue Nozzle Sampler: Refers to the vacuum sampler used to collect dust samples in the HUD National Survey and the Baltimore R&M Pilot study. The sampling flow rate is cited as 16 liters per minute. The sampler consists of a rotary vane pump connected to the same filter and sampling cassette used in the DVM sampler.

Body-Lead Burden: The level of lead carried in a body.

BRM Sampler: Refers to the vacuum sampler developed and utilized to collect dust in EPA's Baltimore Repair and Maintenance Study. It is a modified version of the HVS3 sampler, employing a portable handheld vacuum and other modifications to make it easy to use and to access small areas.

Confidence Interval: An interval that contains the true value of a parameter with a certain degree of confidence.

---

[*] As defined in Section 1001 of Title X and Section 401 of TSCA Title IV amendment.

Conversion Factors: Use of regression models in this risk analysis to convert observed lead measurements from one format to another, typically to correct for differences in dust collection method. For example, a conversion factor was used to express the Blue Nozzle vacuum dust-lead loadings reported in the HUD National Survey as wipe-equivalent dust-lead loadings in order to determine which housing units in the HUD National Survey exceeded example standards for wipe dust-lead loadings.

Cumulative Distribution Function (CDF): For any number x, the CDF $F(x)$ of a random variable X is the probability that the observed value of X will be at most x.

[*] **Deteriorated Paint**: Any interior or exterior paint that is peeling, chipping, chalking or cracking or any paint located on an interior or exterior surface or fixture that is damaged or deteriorated.

Dripline Soil Sample: Any soil sample collected from the drip line area about the residence. This is usually approximately 1-3 feet from the side (e.g. foundation) of the house, under the eaves.

Dry Room: (see Wet Room).

Dust Abatement: Removing settled dust from a housing unit using HEPA vacuums and wet mopping.

Dust Cleaning: Intervention where settled dust that is likely to be lead-contaminated is removed from residential surfaces using HEPA vacuums and wet mopping.

Dust-Lead Loading: Dust-lead loading measures the mass of lead collected per surface area sampled and is usually expressed in terms of micrograms of lead collected per square foot sampled (µg Pb/ft²).

Dust-Lead Concentration: Dust-lead concentration measures the mass of lead collected per mass of dust collected and is usually stated in terms of micrograms of lead collected per gram of dust collected (µg Pb/g dust).

DVM Sampler: A device used to collect dust samples using a vacuum (personal air sampler) operating at a rate of two to three liters of air per minute. It was designed to collect only dust that would most likely stick to a child's hand, not total lead on a surface. Thus, it tends to have low collection efficiency for particles larger than 250 microns.

[*] As define in Section 1001 of Title X and Section 401 of TSCA Title IV amendment.

Efficacy: Refers to the effectiveness of a method of abatement and is defined as the generalized evaluation of several key factors including the usability of a method, its hazard abatement effectiveness, and the amount of hazardous dust lead generated by a method, measured by air and post-cleanup wipe samples.

Empirical Model: A statistical regression model developed for this risk analysis from data collected in the Rochester lead-in-Dust study. The resulting model which predicts geometric mean blood-lead concentration for children aged 12-30 months as a function of environmental lead levels (dust-lead loading, soil-lead concentration, extent of deteriorated lead-based paint hazard) is used to predict a national distribution of children's blood-lead levels for this age group.

Encapsulation: A method of "abatement" that involves the coating and sealing of surfaces with durable coatings formulated to be elastic, long-lasting (e.g., at least 20 years), and resistant to cracking, peeling, algae, and fungus.

Enclosure: The resurfacing or covering of surfaces by sealing or caulking them with mechanically affixed, durable materials so as to prevent or control chalking, flaking, lead-containing substances from being part of house dust or accessible to children.

Entryway Soil: Any soil sample collected immediately adjacent to the entryway of the residence.

EPI Study: A targeted epidemiology study which measures both children's blood-lead concentrations and environmental lead levels as well as other factors (e.g., behavioral, demographic) influencing a child's blood-lead level.

Epidemiology: In broad terms epidemiology is concerned with the distribution of disease, and it is now customary to include within its orbit the study of chronic disease as well as communicable diseases which give rise to epidemics of the classical sort.

Expected Value: The average value of a statistic if it were calculated from an infinite number of equal-sized samples from a given population.

Exposure Route: The manner by which a chemical or pollutant enters an organism after contact (e.g., by ingestion, inhalation).

Exposure Pathway: The physical course a chemical or pollutant takes from its source to the organism exposed.

Exposure: Contact between a chemical, physical, or biological agent (e.g., lead) with the outer boundary of an organism (e.g., a child's skin). Exposure is quantified as the concentration of the agent in the medium in contact integrated over the time duration of that contact.

[*] **Friction Surface**: An interior or exterior surface that is subject to abrasion or friction, including certain window, floor, and stair surfaces.

Geometric Mean: The n^{th} root of the product of n values. Also, the exponentiation of the "arithmetic mean" of a set of n natural log-transformed values.

Geometric Standard Deviation (GSD): The exponentiation of the "standard deviation" of a set of n natural log-transformed values.

HEPA: A High Efficiency Particulate Accumulator vacuum used in dust cleaning, fitted with a filter capable of filtering out particles of 0.3 microns or greater from a body of air at 99.97 percent efficiency or greater.

Histogram: A bar graph associating frequencies or relative frequencies with data intervals. The values of the variable are by convention represented on the horizontal scale, and the vertical scale represents the frequency or relative frequency of data values in each standard grouping of possible values for the variable. It illustrates the general shape of the observed data distribution.

Human Exposure Studies: Studies which investigate the association between elevated blood-lead concentration and elevated levels of lead in a child's residential environment. Examples of human exposure studies are the Rochester Lead-in-Dust study and the Brigham and Women's Hospital Longitudinal study.

HVS3 Vacuum Sampler: Vacuum method originally developed to measure pesticides in house dust and is now recognized as an ASTM standard for collecting floor dust samples to be analyzed for lead content.

IEUBK Model: EPA's Integrated Exposure Uptake Biokinetic Model for Lead, designed to model exposure from lead in air, water, soil, dust, diet, and paint and other sources using pharmacokinetic modeling methods to predict blood-lead concentrations in children 6 months to 7 years of age.

[*] **Impact Surface**: An interior or exterior surface that is subject to damage by repeated impacts, for example, certain parts of door frames.

Individual Risks: Hazards posed for children exposed to specified levels of lead in certain media within the residential environment.

[*] As define in Section 1001 of Title X and Section 401 of TSCA Title IV amendment.

Intelligence Quotient (IQ): A score used to express the apparent relative intelligence of a person determined by dividing his/her mental age as reported on a standardized test by his/her chronological age and multiplying by 100. This risk analysis used IQ score decrement as a means of measuring the neurological effects of lead.

Intercept: See **Slope**.

[*] **Interim Controls**: A set of measures designed to temporarily reduce human exposure or likely human exposure to lead-based paint hazards, including specialized cleaning, repairs, maintenance, painting, temporary containment, ongoing monitoring of lead-based paint hazards or potential hazards, and the establishment and operation of management and resident education programs.

Intervention: A procedure implemented to reduce or eliminate a lead-based paint hazard within a specific medium within a residence, when some type of mechanism is triggered for that medium (e.g., dust-lead standard is exceeded). Interventions considered in this risk analysis include dust cleaning, soil removal, paint maintenance, and paint abatement.

Intervention Studies: Studies which investigate the impact on children's blood-lead concentration of reducing childhood lead exposure via a range of intervention strategies. Intervention studies can contribute to conclusions about whether specific lead exposures are the cause behind elevated blood-lead concentration. Examples of intervention studies are the Baltimore R&M study and the Urban Soil-Lead Abatement Demonstration Project.

***Lead-Based Paint Hazard**: Any condition that causes exposure to lead from lead-contaminated dust, lead-contaminated soil, lead-contaminated paint that is deteriorated or present in accessible surfaces, friction surfaces, or impact surfaces that would result in adverse human health effects as established by EPA.

***Lead-Based Paint (LBP)**: Dried paint film that has a lead content exceeding 1.0 mg/cm² or 0.5 percent (5,000 parts per million (ppm)) by weight.

***Lead-Contaminated Soil**: Bare soil on residential real property that contains lead at or in excess of the levels determined to be hazardous to human health by EPA.

Lead-Contaminated Dust: Surface dust in residential dwellings that contains an area or mass concentration of lead in excess of levels determined by EPA to pose a threat of adverse health effects in pregnant women or young children.

Log-Linear Regression Model: A regression model in which the natural logarithm of the independent (predictor) variables is taken before fitting the model.

---

[*] As define in Section 1001 of Title X and Section 401 of TSCA Title IV amendment.

Lognormal Distribution: A nonnegative random variable X is said to have a lognormal distribution if the natural logarithm of X has a normal (Gaussian) distribution.

Maximum, Minimum and Range: The largest and smallest observations in a data distribution are called *maximum* and *minimum* respectively. The difference between the maximum value and the minimum value is defined as the *range*.

Measurement Error: Error in an observed measurement attributable to sampling, laboratory, spatial and/or temporal variability.

Measurement Error Model: A regression model which attempts to account for measurement error in the observed predictor variables.

Meta-Analysis: The statistical integration of the results of independent studies.

Microgram (µg): A microgram is 1/1,000,000 of a gram or 1/1,000 of a milligram.

Monte Carlo Analysis: An estimation method where approximations are obtained by repeated random sampling or simulation.

Negative Predictive Value (NPV): Probability of a resident child having a blood-lead concentration below some specified threshold value, given that observed lead levels in a specified medium within the dwelling is below the standard for that medium.

90% Confidence Bound on a Statistic: The upper and lower limits of a 90% confidence interval.

Paint Maintenance: Intervention where all surfaces with deteriorated lead-based paint are repaired by feathering the edges of deteriorating paint and repainting with new, lead-free paint.

Paint Abatement: Intervention where all surfaces with deteriorated lead-based paint are encapsulated, enclosed, or removed using currently acceptable practices and materials.

Parameter: A characteristic of a population, such as the population mean or variance.

Percentile: A particular value in a set or distribution of numbers for which a specified percentage of the numbers are less than the given value. For instance, the 5th percentile of a set of blood-lead concentrations is the blood-lead concentration value such that 5% of the numbers are less than the value and 95% are greater than it. The 50th percentile is also known as the median.

Performance Characteristics Analysis: An analysis used to characterize the performance of options for the §403 standards based on the data from IEUBK model or Empirical model.

Perimeter Soil Sample: Any soil sample collected from the perimeter or remote areas of the residence's yard. (Note: in the Rochester Lead-in-Dust study, this terminology referred to samples collected adjacent to the foundation).

Pharmacokinetics: The study of the time course of absorption, distribution, metabolism, and excretion of a foreign substance (e.g., a drug or pollutant) in an organism's body.

Pica: An abnormal tendency to mouth or attempt to consume non-food objects, such as paint chips.

Piecewise Linear Function: The domain of a function divided into finite pieces such that in each piece the function is linear.

Play-yard Soil Sample: Any soil sample collected in areas where the child usually played. In the HUD National Survey, this was frequently a local playground. In other studies, this refers to an exterior site at the residence.

Population: A population of items is defined to be any set of items for which one wants to study and make inferences. Associated with each item in a population are one or more numbers or attributes of interest, which are called variables.

Population Risks: Hazards posed by childhood lead exposure to our nation as a whole.

Positive Predictive Value (PPV): Probability of a resident child having a blood-lead concentration above some specified threshold value given that observed lead levels in a specified medium within the dwelling is above the standard for that medium.

Primary Prevention Intervention: A *primary prevention intervention* prevents human exposure before it occurs (e.g. paint abatement occurs in the home before a new family with children moves in).

Probability Samples: Samples selected from a statistical population such that each sample has a known probability of being selected.

Probability: Given an experiment with an associated sample space, the objective of probability is to assign to each event a number, which will provide a measure of the likelihood that *A* will occur when the experiment is performed.

Random Samples: Samples selected from a statistical population such that each sample has an equal probability of being selected.

[*] **Reduction**: Measures designed to reduce or eliminate human exposure to lead-based paint hazards through methods including interim controls and abatement.

Regression Model: A statistical representation of the relationship between a dependent variable such as blood-lead concentration to one or more independent variables such as environmental lead exposures. For example, a regression model could indicate that blood-lead concentration is an additive function of environmental lead levels.

Removal and Replacement: A method of abatement that entails removing substrates such as windows, doors, trim, or soil that have lead-contaminated surfaces and installing new (and presumably lead-free) or deleaded components.

Residual Error: The difference between the modeled predicted value of a random variable under specified conditions and the observed value of that variable under the same conditions.

Risk: The probability of deleterious health or environmental effects.

Risk Assessment: Within the context of this risk analysis report, risk assessment is that portion of the risk analysis consisting of hazard identification (Chapter 2), exposure assessment (Chapter 3), dose-response assessment (Chapter 4), and risk characterization (Chapter 5). Within the context of identifying lead-based paint hazards in a residence, risk assessment is an on-site investigation to determine and report the existence, nature, severity, and location of lead-based paint hazards within a specific residential dwelling.

Rochester Multimedia Model: A regression model obtained in the process of developing the "empirical model" (using data from the Rochester Lead-in-Dust study) which expresses blood-lead concentration for children aged 12-31 months as a function of environmental-lead levels (dust-lead loading, soil-lead concentration, extent of deteriorated lead-based paint hazard). This model differs from the empirical model in that it does not take into account measurement error in the predictor variables and assumes dust-lead loadings are based on wipe collection techniques. This model was used to characterize individual risks in this risk analysis.

Sample: A small part of something designed to show the nature or quality of the whole. Exposure-related measurements are usually samples of environmental or ambient media, exposures of a small subset of a population for a short time, or biological samples, all for the purpose of inferring the nature and quality of parameters important to evaluating exposure.

[*] As define in Section 1001 of Title X and Section 401 of TSCA Title IV amendment.

Sampling Weights: In a complex survey design, a sampling weight is assigned to a sampling unit to denote the total number of units in the population that is represented by that sampling unit. Sampling weights are necessary to make results of the survey representative of the population. For example, the sampling weight assigned to one of the 284 households in the HUD National Survey represents the number of homes that house represents nationally.

Secondary Prevention Intervention: A *secondary prevention intervention* reduces or eliminates human exposure on behalf of humans already exposed to the targeted hazard (e.g. paint abatement occurs in the home of a child who has an elevated blood-lead concentration).

Sensitivity Analysis: An investigation to determine the extent to which variations in key assumptions and approaches affect the results and conclusions of the analysis.

Sensitivity and Specificity: *Sensitivity* is the probability of a dwelling being above the media standards (e.g., soil lead, dust lead, etc.) given that there is a resident child with blood concentration above some specified threshold value. *Specificity* is the probability of a dwelling being below the media standards given that there is a resident child with blood concentration below some specified threshold value.

Sirchee-Spittler Sampler: Vacuum method used to collect dust samples in the Boston and Baltimore phases of EPA's Urban Soil-Lead Abatement Demonstration Project (USLADP). It is a hand-held, battery-powered vacuum unit designed to collect forensic evidence.

Slope: If the regression model is a simple regression model such that $y=\alpha+\beta x+\mathbf{e}$, then β is called the **slope**, and α is called the **intercept**. The slope is interpreted as the amount by which y changes when x is increased by one unit.

Soil Removal: Intervention where soil from areas with elevated lead concentrations are removed and replaced with clean soil, or the areas are permanently covered.

Soil-Lead Concentration: A measure of the mass of lead collected per mass of soil collected and is usually stated in terms of micrograms of lead collected per gram of soil collected (µg Pb/g soil). These units are also sometimes referred to as parts per million (ppm).

Standard Error: The standard deviation of errors around a fitted regression model.

Standard Deviation: A measure of the dispersion of a set of values that is the square root of the "arithmetic mean" of the squares of the deviation of each value from the "arithmetic mean" of the values.

Subpopulation: A subset of the population of interest that is used for analysis. Usually the subpopulation is taken to be representative of the entire population.

Tails: The portion of a distribution containing extreme values are called the tails of the distribution.

Tap Weight: The weight of the dust that was tapped out of the blue nozzle vacuum cassette and analyzed for lead. Note that a dust sample's tap weight is lower than its actual weight, as some dust may remain in the cassette.

[*]**Target Housing**: Any housing constructed prior to 1978, except for housing of the elderly or persons with disabilities (unless any child who is less than 6 years of age resides or is expected to reside in such housing for the elderly or persons with disabilities), or any 0-bedroom dwelling.

Threshold: The value above which something is true or will take place and below which it is not or will not.

True Negative Rate: Alternative terminology for *specificity*.

True Positive Rate: Alternative terminology for *sensitivity.*

Uptake: The process by which a substance is absorbed into the body.

Vacuum Sample: Collecting dust over a specified area by vacuuming the area. The contents of the vacuum bag or filter cassette are then analyzed for the amount of dust and the amount of lead. Results from vacuum sampling can be expressed as "dust-lead loadings" or "dust-lead concentrations".

Variability: A measure used to describe how data vary about the center of the distribution. It also tells the spread of the data.

Wet Room: An interior room in a house which is either a kitchen, bathroom, laundry, or utility room is classified as a 'wet room', otherwise the room may be classified as a 'dry' room. Terminology used in the HUD National Survey.

Window Sill: The portion of the horizontal window ledge that protrudes into the interior of the room, adjacent to the window sash when the window is closed.

Window Trough: The portion of the horizontal window sill that receives the window sash when the window is closed, often located between the storm window and the interior window sash. This is also sometimes referred to as a **window well**.

[*] As define in Section 1001 of Title X and Section 401 of TSCA Title IV amendment.

Wipe Sample: Dust that is collected over a specified area by wiping the area with a moist cloth. The cloth and the dust on the cloth are then analyzed for the amount of lead. Results from wipe sampling are in the form of "dust-lead loadings." Section 403 standards for lead in dust will likely be specified in terms of wipe dust-lead loadings.

XRF: "X-ray fluorescence" is a principle used by instruments to determine the lead concentration in substances, usually in milligrams of lead per square centimeter of surface area (mg/cm^2).

APPENDIX B

Health Effects Associated with Exposure to Lead and Internal Lead Doses in Humans

Table B-1. Health Effects Associated with Exposure to Lead and Internal Lead Doses in Humans.

Duration of Exposure	System	Effect	Blood Lead Levels at which Effect is Observed (µg/dL)	Reference
< 1 yr (occup)		Increase in death due to hypertension, nephritis, neoplasms	63-80	Cooper et al., 1985, 1988
NS (occup)		Increase in death due to cerebrovascular disease, nephritis, and/or nephrosis	NS	Fanning 1988; Malcolm and Barnett 1982; Michaels et al. 1991
< 3 yr (occup)		No increase in deaths	34-58 (means)	Gerhardsson et al. 1986b
NS		Acute encephalopathy resulting in death in children	125-750	NAS 1972
2 wk - > 1 yr (occup)	Cardiovascular	Increased blood pressure	≥ 30 - 120	deKort et al. 1987; Pollock and Ibels 1986; Marino et al. 1989; Weiss et al. 1986, 1988
> 1 yr (occup)	Cardiovascular	No effect on blood pressure	40 (mean)	Parkinson et al. 1987
> 1 yr (occup)	Cardiovascular	Ischemic electrocardiogram changes	51 (mean)	Kirkby and Gyntelberg 1985
NS (general population)	Cardiovascular	Increased blood pressure	44.9 (mean)	Khera et al. 1980
NS (general population)	Cardiovascular	Increased systolic pressure by 1-2 mmHg and increased diastolic pressure by 1.4 mmHg with every doubling in blood-lead level; effect most prominent in middle-aged white men	7-38	Coate and Fowles 1989; Harlan 1988; Harlan et al. 1988; Landis and Flegal 1988; Pirkle et al. 1985; Schwartz 1988
NS (general population)	Cardiovascular	No significant correlation between blood pressure and blood-lead levels	6-13 (median) or NS	Elwood et al. 1988; Grandjean et al. 1989; Neri et al. 1988; Staessen et al. 1990, 1991
NS (general population)	Cardiovascular	Degenerative changes in myocardium, electrocardiogram abnormalities in children	6-20	Silver and Rodriguez-Torres 1968

Table B-1. Health Effects Associated with Exposure to Lead and Internal Lead Doses in Humans. (Continued)

Duration of Exposure	System	Effect	Blood Lead Levels at which Effect is Observed (µg/dL)	Reference
NS (acute) (occup)	Gastrointestinal	Colic (abdominal pain, constipation, cramps, nausea, vomiting, anorexia, weight loss)	40-200	Awad et al. 1986; Baker et al. 1979; Haenninen et al. 1979; Holness and Nethercott 1988; Kumar et al 1987; Marino et al. 1989; Matte et al. 1989; Muijser et al. 1987; Pagliuca et al. 1990; Pollock and Ibels 1986; Schneitzer et al. 1990
NS (acute) (general population)	Gastrointestinal	Colic in children	60-100	U.S. EPA 1986; NAS 1972
NS (occup)	Hematological	Increased ALAS and/or decreased ALAD	87 or NS (correlated with blood-lead level)	Alessio et al. 1976; Meredith et al. 1978; Wada et al. 1973
NS (general population)	Hematological	Decreased ALAD	3-56 (adult) No threshold (children	Chisholm et al. 1985; Lauwerys et al. 1978; Roels et al. 1976; Roels and Lauwerys 1987; Secchi et al. 1974
NS (occup)	Hematological	Increased urinary or blood ALA	$\prec$ 40-50, 87 (mean) or NS	Lauwerys et al. 1974; Meredith et al. 1978; Pollock and Ibels 1986; Selander and Cramer 1970
NS (general population)	Hematological	Increased urinary ALA	$\succ$ 35 (adult) 25-75 children	NAS 1972; Roels and Lauwerys 1987
NS (general population)	Hematological	Increased FEP	$\geq$ 25-35	Grandjean and Lintrup 1978; Roels et al. 1975
NS (general population)	Hematological	Increased EP	30-40 (males 20-30 (females)	Roels and Lauwerys 1987; Roels et al. 1975, 1976, 1979; Stuick 1974
NS (general population	Hematological	Increased ZPP	$\geq$ 15 (children)	Hammond et al. 1985; Piomelli et al. 1982; Rabinowitz et al. 1986; Roels and Lauwerys 1987; Roels et al. 1976
NS (general population)	Hematological	Increased urinary coproporphyrin	$\geq$ 35 (children) $\geq$ 40 (adults)	U.S. EPA 1986
NS (occup)	Hematological	Decreased hemoglobin with or without basophilic stippling of erythrocytes	$\geq$ 40	Awad et al. 1986; Baker et al. 1979; Grandjean 1979; Lilis et al. 1978; Pagliuca et al. 1990; Tola et al. 1973; Wada et al. 1973

Table B-1. Health Effects Associated with Exposure to Lead and Internal Lead Doses in Humans. (Continued)

Duration of Exposure	System	Effect	Blood Lead Levels at which Effect is Observed (µg/dL)	Reference
NS (general population)	Hematological	Decreased hemoglobin	≥ 40 (children)	Adebonojo 1974; Betts et al. 1973; Pueschel et al. 1972; Rosen et al. 1974
NS (general population)	Hematological	Anemia (hematocrit of < 35%)	> 20 (children)	Schwartz et al. 1990
NS (occup)	Hematological	Decreased Py-5[1]-N	NS	Buc and Kaplan 1978; Paglia et al. 1975, 1977
NS (general population)	Hematological	Decreased Py-5[1]-N	7-80 (children)	Angle and McIntire 1978; Angle et al. 1982
NS (acute) (general population)	Hepatic	Decreased mixed function oxidase activity	NS (children)	Alvares et al. 1975; Saenger et al. 1984
NS (chronic) (occup)	Renal	Chronic Nephropathy	40 - > 100	Biagini et al. 1977; Cramer et al. 1974; Lilis et al. 1968; Maranelli and Apostoli 1987; Ong et al. 1987; Pollock and Ibels 1986; Verschoor et al. 1987; Wedeen et al. 1979
1-30 yr (occup)	Renal	No effect on renal function	40-61	Buchet et al. 1980; Huang et al. 1988
NS (chronic) (general population)	Renal	Renal (impairment with gout or hypertension)	18-26 µg/dL	Batuman et al. 1981, 1983
NS (acute) (general population)	Renal	Aminoaciduria; Fancoi syndrome	> 80 (children)	Chisholm 1962; Pueschel et al. 1972
0.1-20 yr (chronic) (occup)	Other	Decreased thyroxin (T_4)	≥ 56	Tuppurainen et al. 1988
NS (chronic) (general population)	Other	No effect on thyroid function in children	2-77 (levels measured)	Siegel et al. 1989
NS (general population	Other	Negative correlation between blood lead and serum 1,25-dihydroxyvitamin D in children	12-120	Mahaffey et al. 1982; Rosen et al. 1980
NS (chronic) (general population)	Other	No effect on vitamin D metabolism in children	5-24 (levels measured)	Koo et al. 1991

Table B-1. Health Effects Associated with Exposure to Lead and Internal Lead Doses in Humans. (Continued)

Duration of Exposure	System	Effect	Blood Lead Levels at which Effect is Observed (µg/dL)	Reference
NS (chronic) (general population)	Other	Growth retardation in children	≥ 30-60; Tooth lead > 18.7 µg/g	Angle and Kuntzelman 1989; Lauwers et al. 1986; Lyngbye et al. 1987
NS (chronic) (general population)	Other	No association between blood-lead levels and growth in children	10-47 (levels measured)	Greene and Ernhart 1991; Sachs and Moel 1989
< 18 yr (occup)	Immunological	Depression of cellular immune function, but no effect on humoral immune function	21-90	Alomran and Shleamoon 1988; Ewers et al. 1982
NS (acute)	Neurological	Encephalopathy (adults)	50 - > 300	Kehoe 1961; Kumar et al. 1987; Smith et al. 1938
NS (acute and chronic) (occup)	Neurological	Neurological signs and symptoms in adults including malaise, forgetfulness, irritability, lethargy, headache, fatigue, impotence, decreased libido, dizziness, weakness, paresthesia	40-80	Awad et al. 1986; Baker et al. 1979; Campara et al. 1984; Haenninen et al. 1979; Holness and Nethercott 1988; Marino et al. 1989; Matte et al. 1989; Pagliuca et al. 1990; Parkinson et al. 1986; Pasternak et al. 1989; Pollock and Ibels 1986; Schneitzer et al. 1990; Zimmerman-Tansella et al. 1983
NS (occup)	Neurological	Neurobehavioral function in adults; disturbances in oculomotor function, reaction time, visual motor performance, hand dexterity, IQ test and cognitive performance, nervousness, mood, coping ability, memory	40-80	Arnvig et al. 1980; Baker et al. 1983; Baloh et al. 1979; Campara et al. 1984; Glickman et al. 1984; Haenninen et al. 1978; Hogstedt et al. 1983; Mantere et al. 1982; Spivey et al. 1980; Stollery et al. 1989; Valciukas et al. 1978; Williamson and Teo 1986
NS (occup)	Neurological	No effect on neurobehavioral function in adults	40-60 (levels measured)	Milburn et al. 1976; Ryan et al. 1987
NS (occup)	Neurological	Peripheral nerve function in adults; decreased nerve conduction velocity	30- ≥ 70	Araki et al.1980; Muijser et al. 1987; Rosen et al. 1983; Seppalainen et al. 1983; Triebig et al. 1984
NS (occup)	Neurological	No effect on peripheral nerve function	60-80 (levels measured)	Spivey et al. 1980

Table B-1. Health Effects Associated with Exposure to Lead and Internal Lead Doses in Humans. (Continued)

Duration of Exposure	System	Effect	Blood Lead Levels at which Effect is Observed (µg/dL)	Reference
NS (general population)	Neurological	Neurological signs and symptoms in children and encephalopathy	60-450 (effects other than encephalopathy); > 80-800 (encephalopathy)	Bradley and Baumgartner 1958; Bradley et al. 1956; Chisolm 1962, 1965; Chisolm and Harrison 1956; Gant 1938; Rummo et al. 1979; Smith et al. 1983
NS (general population)	Neurological	Neurobehavioral function in children: lower IQS and other neuropsychologic deficits	40-200	dela Burde and Choate 1972, 1975; Ernhart et al. 1981; Kotok 1972; Kotok et al. 1977; Rummo et al. 1979
NS (general population)	Neurological	Neurobehavioral function in children: slightly decreased performance on IQ tests and other measures of neuropsychological function	Tooth lead: 6 - > 30 µg/g Blood lead: 6-60	Bellinger and Needleman 1983; Bergomi et al. 1989; Fulton et al. 1987; Hansen et al. 1989; Hawk et al. 1986; Needleman et al. 1979, 1985, 1990; Schroeder et al. 1985; Schroeder and Hawk 1987; Silva et al. 1988; Wang et al. 1989
NS (general population)	Neurological	No correlation between blood-lead levels and permanent effects on neurobehavioral development in children	10-15	Cooney et al. 1989; Harvey et al. 1984, 1988; Lansdown et al. 1986; McBride et al. 1982; Ernhart and Greene, 1990; Dietrich et al. 1987a; Bellinger et al. 1989a; McMichael et al. 1986; Pocock et al. 1989; Smith et al. 1983; Winneke et al. 1984
NS (general population)	Neurological	Decrease in hearing acuity in children	4-60	Schwartz and Otto 1987
NS (general population)	Neurological	Alterations in peripheral nerve function in children	20-30	Erenberg et al. 1974; Landrigan et al. 1976; Schwartz et al. 1988; Seto and Freeman 1964
prenatal (general population)	Developmental	Decreased growth rate	7.7	Shukla et al. 1989
prenatal (general population)	Developmental	Reduced birth weight and/or reduced gestational age, and/or increased incidence of stillbirth and neonatal death	12-17	Bornschein et al. 1989; McMichael et al. 1986; Moore et al. 1982; Ward et al. 1987; Wibberley et al. 1977

Table B-1. Health Effects Associated with Exposure to Lead and Internal Lead Doses in Humans. (Continued)

Duration of Exposure	System	Effect	Blood Lead Levels at which Effect is Observed (µg/dL)	Reference
NS (general population)	Developmental	No association between blood-lead levels and birth weight, gestational age, or other neonatal size measures	3-55	Greene and Ernhart 1991; Factor-Litvak et al. 1991
NS (general population)	Developmental	Impaired mental development in children	10-15	Baghurst et al. 1987; Bellinger et al. 1984, 1985a, 1985b, 1986a, 1986b, 1987a, 1987b; Bornschein et al. 1989; Dietrich et al. 1986, 1987a, 1987b; Ernhart et al. 1985, 1986, 1987; McMichael et al. 1988; Rothenberg et al. 1989; Wigg et al. 1988; Winneke et al. 1985a, 1985b; Wolf et al. 1985; Vimpani et al. 1985, 1989
NS (general population)	Developmental	Inverse correlation between blood-lead levels and ALA and ALAD activity	10-33 (mean)	Haas et al. 1972; Kuhnert et al. 1977; Lauwerys et al. 1978
NS (general population	Reproductive	Increased incidence of miscarriages and stillbirths in exposed women	≥ 10 or NS	Baghurst et al. 1987; Hu et al. 1991; McMichael et al. 1986; Nordstrom et al. 1979; Wibberley et al. 1977
NS (general population)	Reproductive	No association between blood-lead levels and the incidence of spontaneous abortion in exposed women	2	Murphy et al. 1990
NS (occup)	Reproductive	Adverse effects on testes	40-50	Assennato et al. 1987; Braunstein et al. 1978; Chowdhury et al. 1986; Cullen et al. 1984; Lancranjan et al. 1975; Rodamilans et al. 1988; Wildt et al. 1983

ALA = δ-aminolevulinic acid; ALAD = δ-aminolevulinic acid dehydratase; ALAS = δ-aminolevulinic acid synthase; EP = erythrocyte protoporphyrins; FEP = free erythrocyte protoporphyrins; IQ = intelligence quotient; mmHg = millimeters of mercury; NS = not specified; (occup) = occupational; Py-5^1-N = pyrimidine-5-nucleotidase; wk = week(s); yr = year(s); ZPP = zinc erythrocyte protoporphyrin

APPENDIX C1

Characterizing Baseline Environmental-Lead Levels in the Nation's Housing Stock

APPENDIX C1

CHARACTERIZING BASELINE ENVIRONMENTAL-LEAD LEVELS IN THE NATION'S HOUSING STOCK

As discussed in Section 3.3.1.1, the §403 risk analysis used environmental-lead data from the National Survey of Lead-Based Paint in Housing ("HUD National Survey") to characterize baseline environmental-lead levels in the nation's 1997 housing stock. Here, the term "baseline" refers to conditions prior to implementing interventions in response to §403 rules. Data for 284 privately-owned, occupied housing units included in the HUD National Survey were considered in the characterization. In total, these units represented the entire U.S. privately-owned, occupied housing stock built prior to 1980 (USEPA, 1995a). Due to the complex sampling design employed, the HUD National Survey assigned sampling weights to each unit, which equaled the number of privately-owned, occupied housing units in the national housing stock built prior to 1980 that were represented by the unit (USEPA, 1995g).

In order to use the information from the HUD National Survey to represent baseline environmental-lead levels in the 1997 national housing stock, the following steps were taken:

1. Update the sampling weights assigned in the HUD National Survey to reflect the 1997 housing stock (including publicly-owned units).

2. Determine the total number of children residing in the housing units represented by each sampling weight.

3. Summarize the environmental-lead levels within each surveyed unit.

Methods for conducting each of these steps, and the results from implementing these methods, are summarized in the following sections.

1.0 UPDATING THE NATIONAL SURVEY SAMPLING WEIGHTS

Characterizing the 1997 national housing stock and its distribution of environmental-lead levels involved updating the sampling weights assigned in the HUD National Survey to reflect the 1997 national housing stock. The tasks performed to update these weights were the following:

1. Identify demographic variables that served to group the housing units by their potential for differing environmental-lead levels.

2. Use information within the National Survey weights and the 1993 American Housing Survey to determine total numbers of 1997 housing units within each of these housing groups.

3. Allocate these 1997 totals among the National Survey units within the housing groups.

The methods developed for each of these tasks are presented in the following subsections.

1.1 IDENTIFY SIGNIFICANT FACTORS ASSOCIATED WITH ENVIRONMENTAL-LEAD LEVELS

In updating the sampling weights of the 284 National Survey units, the units were classified into housing groups according to a set of demographic factors found to have a statistically significant influence on environmental-lead levels in the units. Then, the number of 1997 housing units in each group was determined. By grouping the housing units according to these factors, units within the same group had relatively similar distributions of environmental-lead levels, while units in different groups had considerably different distributions.

In determining an appropriate housing grouping, a set of candidate factors was identified, where these factors satisfied three criteria: 1) they would be either important in an economic analysis for §403 rulemaking, or they were likely to be significantly associated with environmental-lead levels; 2) their values for National Survey units existed within the National Survey database; and 3) their values were measured within the 1993 American Housing Survey, a national survey conducted by the Bureau of the Census and the Department of Housing and Urban Development (HUD) to characterize the nation's housing stock (Bureau of the Census and HUD, 1995). Then, a stepwise regression variable selection analysis selected a subset of these factors which explained the largest proportions of house-to-house variability in the following four environmental-lead measurements:

- A mass-weighted arithmetic average floor dust-lead concentration[*] for the unit (i.e., each measurement was weighted by the mass of the sample);
- An area-weighted arithmetic average floor dust-lead loading for the unit (i.e., each measurement was weighted by the square-footage of the sample area);
- A weighted arithmetic average soil-lead concentration for the unit, where results for samples taken from remote locations were weighted twice as much as results for dripline and entryway samples.
- Maximum XRF paint-lead level in the unit (for units containing lead-based paint[**]).

The set of factors included in this analysis are documented in Table C1-1.

---

[*] Prior to calculating the mass-weighted average, dust-lead concentrations were adjusted to reduce bias associated with underestimated sample weights ("low tap weights") reported in the HUD National Survey for dust samples. The adjustment procedure is documented in USEPA, 1996c.

[**] Lead-based paint was considered present in a unit if its predicted maximum XRF value (as determined by statistical modeling techniques within the HUD National Survey) in either the interior or exterior was greater than or equal to 1.0 mg/cm^2.

Table C1-1. Demographic Factors Included in the Stepwise Regression Analysis.

Factor	How the Factor Categorized Housing Units for the Stepwise Regression Analysis
Year the Unit Was Built	Pre-1940; 1940-1959; 1960-1979
Race of Youngest Child	White/Non-Hispanic; Other
Urbanicity Status	City; Suburb/non-metro
Region of Country	Northeast; Midwest; South; West (U.S. Census regions)
Ownership Status	Owner-occupied; renter-occupied
Number of Units in the Bldg.	One unit; more than one unit
Annual Income of Residents	< $30,000; $30,000 or more

The analysis was performed twice on each endpoint: on data for National Survey units containing lead-based paint (LBP) and for units with no LBP. Table C1-2 provides the observed significance levels of each factor considered in the stepwise regression analyses when these levels were below 0.10. Lower significance levels imply a stronger effect on the measurement. The columns in Table C1-2 correspond to separate regression analyses. Across all analyses, the year in which a unit was built (as categorized by pre-1940, 1940-1959, and 1960-1979) had the strongest and most consistent effect on the environmental-lead level (with floor dust-lead concentration an exception). Statistical significance levels for the effect of year built were consistently less than 0.01. While similar significance levels were occasionally observed for other factors in the table, the extent of significance across the environmental-lead measurements was not as consistent for any other factor. Therefore, the year in which the unit was built was the only factor considered in grouping National Survey units for purposes of updating their weights to 1997.

The stepwise regression analysis assumed that the predicted maximum XRF value is an accurate indicator of whether or not a unit contains LBP. Also, those units with no predicted maximum XRF value were assumed not to contain LBP.

1.2 ESTIMATING NUMBERS OF HOUSING UNITS IN 1997 WITHIN YEAR-BUILT CATEGORIES

In this second task, the number of occupied housing units in 1997, both privately- and publicly-owned, was estimated for each of four categories denoting when the unit was built: pre-1940, 1940-1959, 1960-1979, and post-1979. These categories are hereafter referred to as "year-built categories." The results of this task are presented in Table 3-5 within Chapter 3 of this document.

Table C1-2. Demographic Factors Included in Stepwise Regression Analyses, and Significance Levels Associated With These Factors When Less Than 0.10.[1]

Demographic Factors[2]	Units with predicted maximum XRF value less than 1.0 mg/cm² or missing (n = 40)			Units with predicted maximum XRF value at or above 1.0 mg/cm² (n = 221)			
	Floor Dust-Lead Loading	Floor Dust-Lead Conc.[3]	Soil-Lead Conc.	Floor Dust-Lead Loading	Floor Dust-Lead Conc.[3]	Soil-Lead Conc.	Max. Observed XRF Value[4]
Year the Unit Was Built	< 0.01[5]		< 0.01	< 0.01		< 0.01	< 0.01
Race of Youngest Child			0.04				
Urbanicity Status	0.03						
Region of Country							
Ownership Status							
# Units in the Bldg.				0.01	0.01		
Annual Income of Residents							

[1] Column headings for this table identify the environmental-lead measurement being considered in the analysis and the group of National Survey units whose data are included in the analysis. Each column corresponds to a separate regression analysis. The demographic factors included in the regression analyses are included as rows of the table. As the significance level for a demographic factor gets closer to zero, the effect of the factor on the given environmental measurement is considered more highly statistically significant.

[2] See Table C1-1 for definitions of these factors.

[3] This analysis was performed on unadjusted dust-lead concentrations (i.e., no adjustment was made for bias due to underestimated sample weights).

[4] Regression performed on units where the observed maximum XRF value was at least 1.0 mg/cm².

[5] In the regression analysis of floor dust-lead loading in units without LBP, the effect of the year in which the unit was built was statistically significant with a p-value of less than 0.01 (i.e., significance can be concluded at the 0.01 level).

The primary data source for determining the number of units within each year-built category was the 1993 American Housing Survey (AHS) (Bureau of the Census and HUD, 1995). Data from the 1993 AHS provided estimates of the number of housing units in each year-built category in 1993. However, it was of interest to obtain estimates for 1997, not 1993. Therefore, the 1993 estimates were augmented to reflect additions to and removals from the national housing stock from 1994 to 1997. Once the 1997 estimate of the total within each year-built category was obtained, the total was distributed among the National Survey units in the group using information within the National Survey weights. Details on each of these procedures are now provided.

1.2.1 Characterizing the 1993 National Housing Stock

As in the National Survey, each unit in the 1993 AHS was assigned a weight that was interpreted as the number of units in the national housing stock represented by the given unit. Therefore, placing the AHS units among the four year-built categories and summing the weights of the units within each category yielded the estimated number of units in 1993 for each category.

Only occupied housing units in the 1993 AHS (either publicly-owned or privately-owned) were considered in updating to 1997. The definition of an "occupied" unit was one which was occupied by at least one resident who was classified as not having his/her usual residence elsewhere. Data for 40,931 occupied housing units were available from the 1993 AHS.

1.2.2 Updating the 1993 Housing Stock to 1997

Once the number of housing units in 1993 was determined for each of the four year-built categories, these totals were updated to reflect the 1997 housing stock. Updating the 1993 totals to 1997 was done in the following way:

1. For the post-1979 category, the total number of housing units constructed from 1994 to 1997 and occupied in 1997 was estimated and added to the 1993 total.

2. For all four year-built categories, the total number of housing units occupied in 1993 and lost from the housing stock from 1994 to 1997 was estimated and subtracted from the 1993 total.

In the first step, numbers of new, privately-owned housing units completed in 1994 and 1995 were obtained from Bureau of the Census and HUD (1996). This publication reported estimates of 1,346,900 such units completed in 1994 and 1,311,300 units in 1995. For this analysis, the 1995 estimate was also used in estimating totals for both 1996 and 1997. Therefore, the 1993 estimate for the post-1979 housing category was incremented by 1,346,900 + 3*1,311,300 = 5,280,800 units. Note that this approach assumes that new housing units are completed and occupied within the same year. In addition, no provision was considered for adding new publicly-owned units.

The second step, subtracting the number of housing units occupied in 1993 and lost from the housing stock from 1994 to 1997 within each of the four year-built categories, was more complex. Information on losses was not available by considering only the 1993 AHS. To obtain such information, the 1989 and 1991 AHS databases were obtained. As the AHS retains the same units from survey to survey, it was possible to determine those units that were occupied in one survey and lost from the housing stock by the next. Units were considered lost from the housing stock in a given survey if they were labeled as a "Type C non-interview" in the survey, meaning the unit no longer exists and is dropped from consideration for future surveys. Such losses include demolition, disaster loss, abandoned permit, or the unit was merged with another unit. While moving a house or mobile home from the site also labels the unit as a Type C noninterview, such an instance was <u>not</u> labeled as a loss from the housing stock for this effort, as it is assumed that the unit remains habitable in its new location.

As the AHS is conducted every two years, the probability that a unit is lost from the housing stock over a two-year period was initially estimated from the AHS data. In this procedure, a dataset of information on occupied housing units present in the 1989 AHS was created, with each unit identified by its approximate age in 1989 (in years), by its 1989 sample weight, and by whether or not it was classified as lost from the housing stock in the 1991 AHS.

Similarly, a dataset of information on occupied housing units present in the 1991 AHS was created, with each unit identified by its approximate age in 1991, by its 1991 sample weight, and by whether or not it was classified as lost from the housing stock in the 1993 AHS. Both datasets were combined into a single dataset (without regard to survey year), and a logistic regression analysis was fitted to the combined data to predict the probability of a loss over a two-year period as a function of age (in years). Each data point in the regression analysis was weighted by its sample weight. The resulting prediction model was

$$\text{P[loss over a two-year period]} = \frac{1}{1 + e^{5.82 - 0.0094 * \text{age}}} \qquad \textbf{(1)}$$

where "age" is the age of the unit in years. The probability for a one-year period was roughly one-half of the probability for the two-year period. Table C1-3 provides the predicted probabilities of losses over a one-year period for every five years of age.

Table C1-3. Estimated Probability of an Occupied Housing Unit Becoming Lost from the Housing Stock Over a One-Year Period, Given the Age of Unit.

Age of Unit (yrs)	Probability of Loss	Age of Unit (yrs)	Probability of Loss
5	0.0013	45	0.0023
10	0.0014	50	0.0025
15	0.0015	55	0.0026
20	0.0016	60	0.0028
25	0.0017	65	0.0031
30	0.0018	70	0.0033
35	0.0020	75	0.0036
40	0.0021	80	0.0038

Note: These probabilities were estimated from equation (1) and adjusted to cover a one-year period.

Table C1-4 illustrates how losses from the housing stock from 1993 to 1997 were characterized within each of the four year-built categories considered in the risk analysis. First, an age (in years) associated with each of the four year-built categories was determined for 1993 and 1995. For the 1940-1959, 1960-1979, and post-1979 categories, this age corresponded to the age of a unit built in the middle year of the category. The single age assigned to all units in the pre-1940 category was equal to the age of a unit built in 1939. Then, the probability of loss from 1993-1995 and from 1995-1997 was determined from equation (1) based on the age of the unit; these probabilities are labeled in Table C1-4 as $p_{1993\text{-}95}$ and $p_{1995\text{-}97}$, respectively. The total number of units in the category in 1993 was then reduced by multiplying the total by the product $(1\text{-}p_{1993\text{-}95})*(1\text{-}p_{1995\text{-}97})$ (i.e, the last column of Table C1-4).

Table C1-4. Determining Losses from the Housing Stock from 1993-1997.

Year-Built Category	Age of units in 1993 (yrs.)[1]	Prob. of loss from 1993-1995 ($p_{1993-95}$)[2]	Age of units in 1995 (yrs.)[1]	Prob. of loss from 1995-1997 ($p_{1995-97}$)[2]	Proportion of 1993 Total That Remains in 1997 [3]
Pre-1940	54	0.0052	56	0.0054	0.989
1940-1959	44	0.0045	46	0.0046	0.991
1960-1979	24	0.0034	26	0.0034	0.993
Post-1979	7	0.0026	9	0.0027	0.995

[1] A single age is assigned to all units in a given category according to the approach indicated in the text.
[2] Determined from equation (1).
[3] Equal to $(1-p_{1993-95})*(1-p_{1995-97})$

Besides additions and removals, changes in the number of occupied homes in the national housing stock from 1993 to 1997 are also affected by the number of units that are occupied in 1993 and vacant in 1997, as well as by the number of units that are vacant in 1993 and occupied in 1997. However, in this approach, it was assumed that the number of occupied units in 1993 that become vacant in 1997 was approximately equal to the number of vacant units in 1993 that become occupied in 1997, thereby canceling each other out.

1.3 DETERMINING THE NUMBER OF 1997 UNITS REPRESENTED BY EACH NATIONAL SURVEY UNIT

The procedures outlined in the previous subsection provide a method for estimating total numbers of housing units in 1997 within each of the four year-built categories. The results are displayed in Table 3-5 in Chapter 3 of this report. The housing units were grouped within year-built categories to facilitate the linking of numbers of units with estimated environmental-lead levels. The linking process consisted of classifying the National Survey units among the four categories, then distributing the 1997 total among the National Survey units within each category. This distribution yielded an updated weight for each National Survey unit, reflecting changes in the numbers of units in the year-built category from the time the National Survey was conducted to 1997. A unit's updated weight represented the number of units in the 1997 housing stock associated with the National Survey unit (and therefore with its environmental-lead levels).

The 1997 totals include both privately-owned and publicly-owned housing units, while the 284 National Survey units were privately-owned. Therefore, the revised 1997 weights for the National Survey units represent publicly-owned as well as privately-owned units.

1.3.1 Updating the Weights to Reflect the Pre-1980 Housing Stock

To update the sampling weights for the 284 National Survey units to reflect the pre-1980 housing stock, the units were grouped according to the three pre-1980 year-built categories. (Recall that all National Survey units were built prior to 1980). For these three categories, the updated 1997 weight for each unit in the category was calculated as follows:

$$\text{1997 weight} = (\text{National Survey weight}) * (\text{Updating factor for the category}) \quad \textbf{(2)}$$

where the updating factor was determined as follows:

$$\text{Updating factor} = \frac{\text{\# units in the category in 1997}}{\text{Total National Survey weights in the category}} \quad \textbf{(3)}$$

(The sampling weights assigned in the National Survey were determined according to when the unit was built, whether the unit existed in a single- or multiple-unit building, the Census region in which the unit was located, and whether or not a child less than aged seven years resided in the unit).

Table C1-5 contains the updating factors applied to the National Survey units according to year-built category. As an example, Table C1-5 indicates that the updated 1997 weight for each of the 77 National Survey units in the pre-1940 category equaled the weight assigned in the National Survey multiplied by 0.936.

Table C1-5. Number of National Survey Units in the Pre-1980 Year-Built Categories, and the Multiplicative Factor Used to Update National Survey Weights to 1997.

Year-Built Category	# National Survey Units	Sum of National Survey Weights	Updating Factor
Pre-1940	77	21,020,019	0.936
1940-1959	87	20,472,997	0.963
1960-1979	120	35,686,004	0.980

1.3.2 Updating the Weights to Reflect the Post-1979 Housing Stock

Despite the fact that no HUD National Survey units were built after 1979, it was of interest to use the HUD National Survey data to characterize the entire occupied national housing stock, including those units built after 1979. Therefore, methods were developed to determine how to use environmental-lead information from the HUD National Survey to represent the post-1979 occupied housing stock.

As the post-1979 housing stock was built after the Consumer Product Safety Commission's 1978 ban on the sale of LBP and its use in residences, the post-1979 housing stock

was assumed to be free of LBP. This same assumption was made in the HUD National Survey and is the reason for not including post-1979 housing in the survey. Therefore, only National Survey units not containing LBP were considered in representing post-1979 housing.

To determine whether the entire set of National Survey units without LBP should be considered in representing post-1979 housing or only a subset of these units, data on dust-lead and soil-lead concentrations for units having maximum and minimum XRF measurements below 0.7 mg/cm² were investigated. As the top two plots in Figure C1-1 illustrate, a noticeable relationship exists between lead concentrations and the age of the unit, with higher concentrations associated with older units. In contrast, the bottom two plots in Figure C1-1 show less of a relationship between concentration and age of unit when only units built from 1960-1979 were considered. This finding suggests that older units may be free of LBP, but dust and soil are more likely to remain contaminated with lead than for newer units, either due to previous renovation work on the units or from outside contamination.

As a result of the conclusions made from Figure C1-1, only the 28 National Survey units built between 1960 and 1979 and containing no LBP (predicted maximum XRF measurement less than 1.0 mg/cm²) were selected to represent the post-1979 housing stock. As a result, it was assumed that the environmental-lead levels for these 28 units represented levels that exist in the post-1979 housing stock. These units also were included among those representing the 1960-1979 housing stock. Therefore, the total 1997 sampling weight for these 28 units consisted of two parts: that representing the 1960-1979 housing stock, and that representing the post-1979 housing stock, 1997 weight = (1960-1979 housing stock weight) + (post-1979 housing stock weight), where the 1960-1979 housing stock weight was calculated as described above. The portion representing the post-1979 housing stock was determined by dividing the total number of post-1979 units in 1997 by 28,

$$\text{post-1979 housing stock weight} = (\text{total \# of post-1979 units}) / 28. \quad \textbf{(4)}$$

2.0 POPULATING HOUSING UNITS WITH CHILDREN

To characterize risk reduction that may result from performing interventions in response to §403 rules, it was necessary to estimate numbers of children of specific age groups who reside within the national housing stock. This section documents the methods for populating the 1997 national housing stock with children.

Section 1.0 of this appendix presented methods to revising the sampling weights for HUD National Survey units to reflect the 1997 national housing stock of occupied units. Therefore, each weight represents a subset of the national housing stock. It was desired to link numbers of children with each weight. Two age groups of children were of interest:

- Children aged 12 to 35 months (1 to 2 years)
- Children aged 12 to 71 months (1 to 5 years)

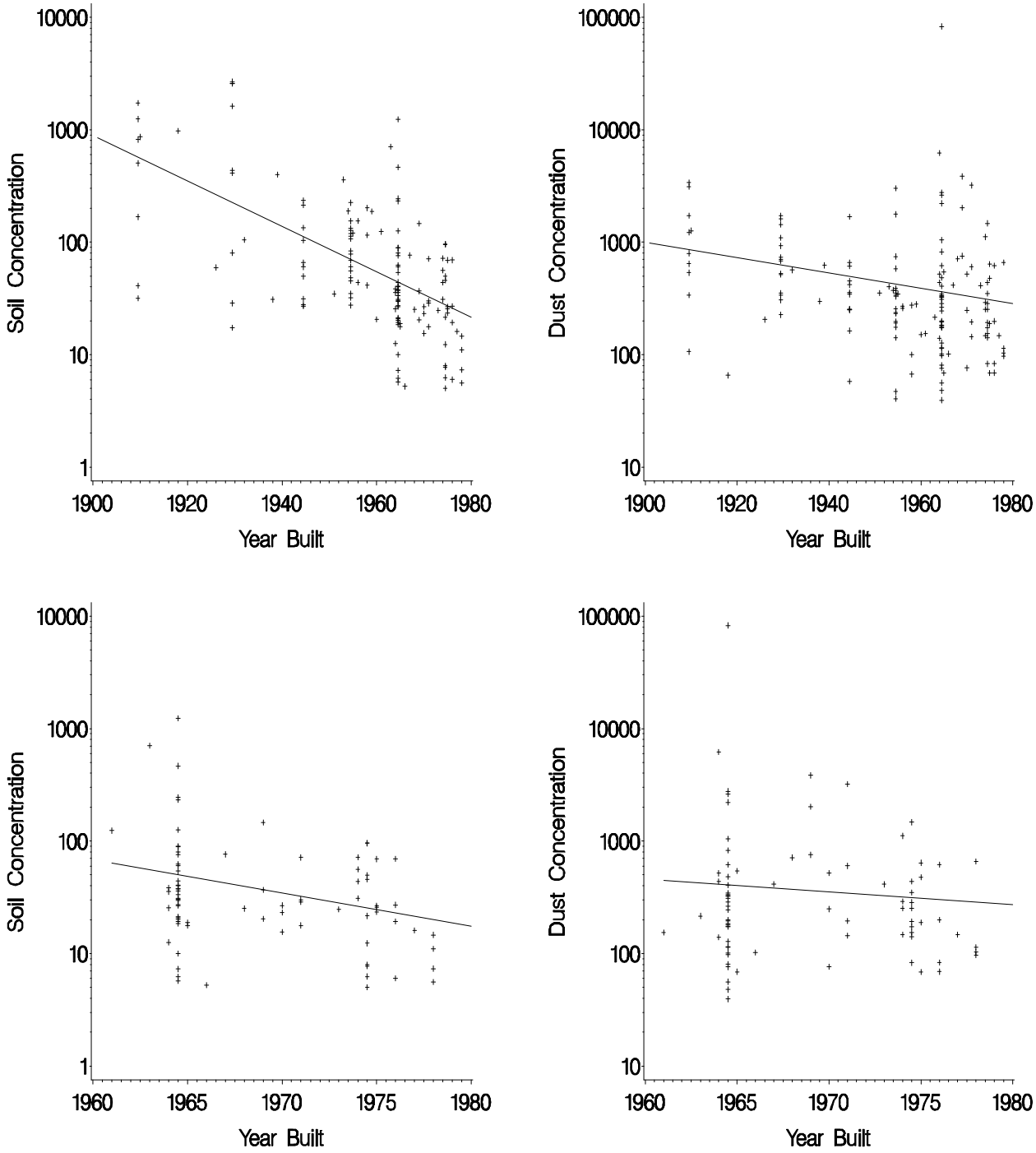


Figure C1-1. Plots of Dust- and Soil-Lead Concentration (µg/g) Versus Age of Unit, for HUD National Survey Units With Maximum XRF Value Less Than 0.7 mg/cm²

The 1-2 year age group was the primary group of interest in this risk analysis, while the 1-5 year age group was considered in the sensitivity analysis within Chapter 5 (Section 5.4.1).

For a given age group of children, the estimated number of children associated with the units represented within a 1997 sampling weight was the product of three statistics:

$$\text{\# children} = (\text{1997 weight}) * (\text{Average \# residents per unit}) * (\text{\# children per person}) \quad \textbf{(5)}$$

As the 1997 weight was determined for each National Survey unit using the methods in Section 1.0 of this appendix, it was necessary to obtain estimates for the latter two statistics in equation (5).

The factor "average # residents per unit" in equation (5) was calculated for the housing group based on information obtained in the 1993 AHS. The 1993 AHS database provided information on up to 15 residents within each housing unit in the AHS. Once these units were placed within the four year-built categories, the average number of people residing in a unit (regardless of age) was calculated for each group. This average ranged from 2.5 to 2.7 across the four year-built categories. A common average of 2.7 residents per unit was used for all units in the national housing stock. While this average was based on 1993 data, it is assumed to also hold for the 1997 housing stock.

The third factor in equation (5), "# children per person," represented the average number of children (of the given age group) per person residing in units within the housing group. This factor was calculated from information presented in Day (1993). This document provided two types of information necessary to calculate average number of children per person:

1. Predicted numbers of births per 1,000 people in the general population within selected years from 1993 to 2050

2. Predicted numbers of people in the general population of specific ages for these selected years.

For 1997, Day (1993) predicted a total of 14.8 births predicted per 1,000 people in the U.S.[*] Therefore, it was assumed that in any subset of occupied housing in 1997, the units within this subset will contain 14.8 children less than one year of age for every 1000 residents.

Day (1993) also provided a predicted number of children of various age groups in the nation in 1997. A total of 3,907,000 children aged 0-11 months, 7,835,000 children aged 12 to

[*] This is a "middle series assumption" birth rate, indicating the level at which assumptions are placed on fertility, life expectancy, and yearly net immigration.

35 months, and 20,066,000 children aged 12 to 71 months were predicted. By dividing each of these latter two statistics by 3,907,000, approximately 2.01 children aged 12 to 35 months and 5.14 children aged 12 to 71 months are predicted in 1997 for every child aged 0-11 months. Thus, using the birth rate in the previous paragraph, a total of 2.01 x 14.8 = 29.7 children aged 12 to 35 months, and 5.14 x 14.8 = 76.1 children aged 12 to 71 months, are predicted in 1997 per 1000 people in the U.S.

Table C1-6 contains estimates of average number of children per unit in the 1997 national housing stock, according to age group. These numbers are the product of the final two factors in equation (5). Therefore, these numbers are multiplied by the 1997 sampling weights for each National Survey unit to obtain an estimated number of children residing in units represented within the weight. By summing the estimates across National Survey units, the total number of children aged 12-35 months and 12-71 months residing within the 1997 national housing stock is obtained by year-built category and for the nation. These results are presented in Table 3-35 in Chapter 3 of this report.

Table C1-6. Estimated Average Number of Children Per Unit in the 1997 National Housing Stock, by Age of Child.

Age Group	Estimated Average Number of Children Per Unit
12-35 months	2.7*0.0297 = 0.080
12-71 months	2.7*0.0761 = 0.205

3.0 SUMMARIZING ENVIRONMENTAL-LEAD LEVELS WITHIN THE HUD NATIONAL SURVEY UNITS

The methods of Sections 1.0 and 2.0 of this appendix were used to link each of the 284 units in the HUD National Survey with an estimated number of units in the 1997 national housing stock and an estimated number of children residing within these units. In this final step, it is necessary to summarize the environmental-lead levels within each National Survey unit.

The following statistics were calculated for each National Survey unit, summarizing the unit's dust-lead loadings and dust-lead concentrations from floors and window sills, and soil-lead concentrations:

- A mass-weighted arithmetic average floor dust-lead concentration* for the unit (i.e., each measurement is weighted by the mass of the sample);

- An area-weighted arithmetic average floor dust-lead loading for the unit (i.e., each measurement is weighted by the square-footage of the sample area);

- A mass-weighted arithmetic average window sill dust-lead concentration* for the unit (i.e., each measurement is weighted by the mass of the sample);

- An area-weighted arithmetic average window sill dust-lead loading for the unit (i.e., each measurement is weighted by the square-footage of the sample area);

- A weighted arithmetic average soil-lead concentration for the unit, where results for samples taken from remote locations were weighted twice as much as results for dripline and entryway samples. If a unit has no soil-lead results for a particular location, the arithmetic average was unweighted (i.e., results for the remaining locations were not weighted).

- An unweighted arithmetic average soil-lead concentration, considering only the dripline and entryway samples for the unit.

- The maximum paint-lead concentration in the interior and the exterior of the unit, as measured by XRF techniques in selected rooms and on selected components within these rooms.

- The amount of damaged lead-based paint measured in the interior and the exterior of the unit.

These summary values were used in the statistical models to represent environmental-lead levels in the national housing stock, in determining health benefits associated with intervention.

In the HUD National Survey database, some units have unrecorded (or "missing") values for dust-lead loadings or concentrations, or soil-lead concentrations, preventing values for one or more of the first six summary statistics above from being calculated. As the values of certain statistics were used as input to the IEUBK and empirical models to predict any risk reductions that may result from performing interventions in response to §403 rules, it was necessary that every housing unit have values for these statistics, even if no data existed for a particular unit. Therefore, an imputation scheme was devised to obtain summary values for units having no data in the National Survey database for the given parameter. In this approach, if a unit did not have data to allow the value of a summary statistic from being calculated, the value assigned to the unit equaled the weighted arithmetic average of those values for units within the same year-built

* Prior to calculating the mass-weighted average, dust-lead concentrations on floors and window sills were adjusted to reduce bias associated with underestimated sample weights ("low tap weights") reported in the National Survey for dust samples.

category and having the same indicator for the presence of LBP, with each value weighted by the 1997 weight for the respective unit. For example, a total of eight National Survey units were built prior to 1940 and contained no LBP. If one of these units had no floor dust-lead loadings, then the summary value of floor-dust-lead loading for this unit would equal the weighted average of the summary values across the other seven units. The inputed values are documented in Table 3-14 of Chapter 3.

Table C1-7 contains a listing of National Survey units within the three year-built categories in which they are classified. Also note that the 28 National Survey units built from 1960-1979 and containing no LBP were listed within a fourth category within Table C1-7, representing the national housing stock built after 1979. The dust-lead concentrations summarized in Table C1-7 were initially adjusted for underestimated sample weights (USEPA, 1996c). Also, dust-lead loadings summarized in Table C1-7 were initially adjusted to reflect loadings that would be obtained if wipe collection techniques were used, rather than the Blue Nozzle vacuum method employed in the HUD National Survey. The method to converting from Blue Nozzle vacuum to wipe loadings is presented in Chapter 4.

Table C1-7 also contains the updated 1997 sampling weights for each unit (as calculated in Section 1.0 of this appendix) and the estimated numbers of children aged 12-35 months and 12-71 months that reside within the units (as calculated in Section 2.0 of this appendix). For the 28 units listed in both the 1960-1979 and post-1979 categories, the sampling weights and numbers of children are only that portion representing units within the category.

Table C1-7. Estimated Environmental Lead Levels in the 1997 Housing Stock, As Determined from National Survey Units

Year Built	National Survey ID	LBP Present?	Wipe Floor Dust-Lead Loading (ug/ft2)	Vac. Floor Dust-Lead Loading (ug/ft2)	BN Floor Dust-Lead Conc. (ug/g)	Wipe W. Sill Dust-Lead Loading (ug/ft2)	Vac. W. Sill Dust-Lead Loading (ug/ft2)	Yardwide Avg. Soil-Lead Conc. (ug/g)	Obs. Max. Interior XRF (mg/cm2)	Obs. Max. Exterior XRF (mg/cm2)	Damaged Interior LBP (ft2)	Damaged Exterior LBP (ft2)	1997 Weight	# Children 12-35 mo.	# Children 12-71 mo.
<1940	0320408	No	8.43	1.72	320.	1.83	0.62	36.5	0.60	--	0.0	0.0	183,864	14,744	37,779
	0320507	No	17.2	4.19	338.	35.6	7.78	113.	--	--	0.0	0.0	183,864	14,744	37,779
	1210806	No	23.1	6.28	970.	17.6	4.55	279.	--	--	0.0	--	121,752	9,763	25,016
	1921709	No	23.5	5.86	448.	35.7	8.28	305.	--	--	0.0	--	199,528	16,000	40,997
	1932300	No	106.	40.0	412.	1220.	166.	279.	--	--	0.0	0.0	199,528	16,000	40,997
	1942606	No	31.4	9.10	246.	2250.	277.	259.	0.60	0.00	0.0	0.0	199,528	16,000	40,997
	1953009	No	0.99	0.13	103.	1.36	0.52	279.	0.60	--	0.0	0.0	199,528	16,000	40,997
	2022507	No	93.3	34.0	589.	176.	31.9	326.	0.60	--	0.0	0.0	1,140,935	91,492	234,428
	0211102	Yes	17.3	4.64	778.	0.14	0.08	84.2	2.8	8.7	0.0	0.0	183,864	14,744	37,779
	0221101	Yes	2.09	0.32	148.	0.80	0.33	394.	0.60	5.1	0.0	0.0	95,766	7,679	19,677
	0221507	Yes	26.7	6.95	975.	449.	65.1	2020.	10.	6.0	0.0	4.8	183,864	14,744	37,779
	0310102	Yes	13.2	3.08	297.	2.58	0.89	138.	0.60	0.60	0.0	0.0	183,864	14,744	37,779
	0310607	Yes	2.83	0.46	63.8	440.	68.3	1240.	3.4	--	0.0	--	183,864	14,744	37,779
	0310706	Yes	2.83	0.48	197.	59.6	12.8	534.	7.1	14.	0.0	0.0	95,766	7,679	19,677
	0311100	Yes	96.6	41.7	1600.	3.03	1.02	711.	5.3	5.8	0.0	57.6	183,864	14,744	37,779
	0320705	Yes	6.40	1.28	406.	0.86	0.35	274.	0.70	27.	0.0	0.0	183,864	14,744	37,779
	0350801	Yes	22.6	6.55	2110.	29.6	6.47	25.9	--	--	0.0	0.0	95,766	7,679	19,677
	0411207	Yes	236.	118.	1810.	14.6	3.88	805.	0.40	0.40	0.0	0.0	244,799	19,630	50,299
	0520106	Yes	4.61	0.81	86.6	4.92	1.54	59.6	0.60	0.40	0.0	0.0	244,799	19,630	50,299
	0520403	Yes	130.	51.8	299.	246.	41.8	102.	0.70	1.8	0.0	0.0	114,632	9,192	23,553
	0520700	Yes	75.7	27.2	938.	6190.	592.	258.	0.60	2.8	0.0	0.0	199,528	16,000	40,997
	0520908	Yes	12.2	2.93	631.	108.	21.2	17.4	0.70	0.60	0.0	0.0	114,632	9,192	23,553
	0711002	Yes	24.6	7.08	537.	8.32	2.41	642.	0.20	13.	0.0	0.0	111,365	8,930	22,882
	0720300	Yes	13.3	3.07	340.	2540.	307.	1460.	12.	0.60	0.0	0.0	111,365	8,930	22,882
	0720706	Yes	16.1	3.80	526.	298.	46.3	841.	8.0	5.0	0.0	24.6	111,365	8,930	22,882
	0721001	Yes	31.0	8.56	326.	2300. *	207. *	80.4	3.3	0.60	0.0	0.0	60,761	4,872	12,485
	0730606	Yes	49.3	14.8	527.	43700.	3150.	372.	10.	8.8	9.4	28.0	111,365	8,930	22,882
	0820506	Yes	6.83	1.32	130.	13.3	3.59	835.	0.70	3.6	0.0	226.8	111,365	8,930	22,882
	0911800	Yes	2.83	0.43	92.2	97.2	19.3	49.8	0.60	0.80	0.0	0.0	111,365	8,930	22,882
	0920900	Yes	4.73	0.84	187.	1.28	0.49	162.	0.60	54.	0.0	0.0	111,365	8,930	22,882
	0941005	Yes	4.84	0.86	244.	896.	127.	1620.	0.80	3.8	0.0	0.0	773,094	61,994	158,848
	0950402	Yes	17.7	4.26	641.	2310.	262.	2000.	0.30	0.30	0.0	0.0	773,094	61,994	158,848
	0951004	Yes	7.52	1.57	522.	101.	18.5	1170.	0.60	6.5	0.0	457.3	773,094	61,994	158,848
	1010909	Yes	23.2	6.63	1240.	24.4	5.84	851.	10.	51.	0.0	0.0	244,799	19,630	50,299
	1011303	Yes	44.2	13.3	1100.	2300. *	207. *	717.	0.80	--	0.0	--	244,799	19,630	50,299
	1011501	Yes	19.0	4.51	616.	48.3	9.95	4620.	0.40	38.	0.0	0.0	114,632	9,192	23,553
	1011600	Yes	46.2 *	17.9 *	451.	2300. *	207. *	392.	0.30	29.	0.0	182.0	244,799	19,630	50,299
	1041607	Yes	2.85	0.45	0.09	1.12	0.44	39.5	0.30	0.30	0.0	0.0	244,799	19,630	50,299
	1221902	Yes	100.	41.4	6320.	14600.	1200.	444.	6.4	11.	0.0	8.4	1,140,935	91,492	234,428
	1250406	Yes	32.8	9.01	2260.	96.4	19.1	628.	6.2	4.9	0.0	0.0	121,752	9,763	25,016
	1251107	Yes	74.1	27.2	1760.	85.7	17.4	1030.	5.0	0.00	0.0	0.0	121,752	9,763	25,016
	1251404	Yes	11.0	2.40	638.	36.3	8.08	569.	20.	4.0	0.0	0.0	1,140,935	91,492	234,428
	1352608	Yes	173.	78.1	2070.	2300. *	207. *	679.	7.0	10.	0.0	141.4	111,365	8,930	22,882
	1353705	Yes	4.85	0.86	451.	7.54	2.17	109.	13.	1.8	11.5	0.0	111,365	8,930	22,882
	1411909	Yes	197.	102.	4340.	7.05	2.06	586.	0.60	7.9	0.0	0.0	95,766	7,679	19,677
	1531201	Yes	134.	53.4	831.	542.	83.0	251.	0.90	14.	0.0	585.7	773,094	61,994	158,848
	1531300	Yes	16.8	4.29	303.	35.0	8.11	105.	3.3	4.4	0.0	112.0	111,365	8,930	22,882
	1631209	Yes	12.9	3.09	215.	210.	37.1	841.	1.4	1.6	0.0	0.0	199,528	16,000	40,997
	1631308	Yes	6.28	1.15	122.	229.	40.0	539.	1.2	1.6	0.0	0.0	199,528	16,000	40,997
	1740901	Yes	81.1	27.7	860.	2300. *	207. *	137.	9.4	15.	89.8	0.0	121,752	9,763	25,016
	1751304	Yes	14.8	3.54	198.	0.02	0.01	358.	2.9	9.5	17.6	0.0	121,752	9,763	25,016
	1820802	Yes	3.96	0.69	105.	0.81	0.32	1430.	0.60	--	0.0	0.0	60,761	4,872	12,485
	1830801	Yes	9.48	1.90	271.	3.07	1.03	841.	6.6	--	18.7	--	60,761	4,872	12,485
	1830900	Yes	375.	194.	3630.	8.85	2.36	841.	4.7	--	0.0	--	199,528	16,000	40,997
	1840503	Yes	114.	47.3	1970.	22.1	5.17	841.	1.2	--	0.0	--	60,761	4,872	12,485
	1851104	Yes	32.8	10.3	316.	414.	66.0	383.	0.60	4.6	0.0	0.0	121,752	9,763	25,016
	1931906	Yes	17.0	4.19	193.	1030.	132.	841.	4.4	2.7	0.8	0.0	199,528	16,000	40,997
	1951904	Yes	44.9	13.1	625.	303.	50.2	841.	6.0	--	0.0	0.0	60,761	4,872	12,485
	1952506	Yes	12.6	2.76	328.	0.38	0.17	841.	1.9	2.4	0.0	0.0	60,761	4,872	12,485
	2121507	Yes	27.2	7.68	281.	2300. *	207. *	860.	1.7	7.1	6.2	25.1	199,528	16,000	40,997
	2240406	Yes	225.	97.7	781.	28400.	2190.	335.	0.60	3.5	0.0	604.8	244,799	19,630	50,299

Table C1-7. Estimated Environmental Lead Levels in the 1997 Housing Stock, As Determined from National Survey Units. (Continued)

Year Built	National Survey ID	LBP Present?	Wipe Floor Dust-Lead Loading (ug/ft2)	Vac. Floor Dust-Lead Loading (ug/ft2)	BN Floor Dust-Lead Conc. (ug/g)	Wipe W. Sill Dust-Lead Loading (ug/ft2)	Vac. W. Sill Dust-Lead Loading (ug/ft2)	Yardwide Avg. Soil-Lead Conc. (ug/g)	Obs. Max. Interior XRF (mg/cm2)	Obs. Max. Exterior XRF (mg/cm2)	Damaged Interior LBP (ft2)	Damaged Exterior LBP (ft2)	1997 Weight	# Children 12-35 mo.	# Children 12-71 mo.
<1940 (cont.)	2311108	Yes	76.3	26.4	1280.	1850.	216.	841.	8.6	0.70	21.9	0.0	1,140,935	91,492	234,428
	2343002	Yes	6.35	1.19	277.	85.6	16.0	256.	2.3	5.7	0.5	1.7	121,752	9,763	25,016
	2410801	Yes	8.95	1.85	612.	67.1	14.1	290.	5.9	7.6	238.6	77.3	121,752	9,763	25,016
	2441608	Yes	5.11	0.98	342.	469.	67.5	609.	9.4	3.9	0.0	0.0	121,752	9,763	25,016
	2521300	Yes	15.3	3.51	150.	1.13	0.42	35.0	0.50	0.60	0.0	0.0	244,799	19,630	50,299
	2541209	Yes	32.8	10.5	161.	254.	42.7	28.6	1.5	0.50	7.0	0.0	114,632	9,192	23,553
	2542009	Yes	11.1	2.31	277.	442.	68.6	125.	8.2	0.90	139.9	0.0	114,632	9,192	23,553
	2550309	Yes	1.32	0.17	57.5	21.8	5.29	76.4	0.60	6.6	0.0	0.0	244,799	19,630	50,299
	2551802	Yes	12.2	3.19	142.	16.7	4.02	159.	1.3	0.50	0.0	0.0	244,799	19,630	50,299
	2651800	Yes	25.1	6.46	399.	5.37	1.66	47.4	0.60	0.30	0.0	0.0	244,799	19,630	50,299
	2710101	Yes	8.47	1.69	261.	1.11	0.44	613.	2.6	5.0	0.0	0.0	244,799	19,630	50,299
	2721009	Yes	17.0	3.93	316.	1020.	142.	110.	2.9	0.50	0.0	0.0	244,799	19,630	50,299
	2931608	Yes	80.6	31.3	813.	401.	64.3	1160.	3.9	7.7	28.8	0.0	183,864	14,744	37,779
	3011103	Yes	79.7	27.8	1310.	808.	114.	1500.	12.	6.9	0.9	0.0	111,365	8,930	22,882
	3011905	Yes	8.84	1.98	764.	1130.	129.	2750.	10.	3.3	6.6	0.0	773,094	61,994	158,848
	3020401	Yes	8.94	1.79	327.	198.	31.9	1390.	0.60	5.3	0.0	16.5	773,094	61,994	158,848
													19,676,320	1,577,844	4,042,893
1940-1959	0340406	No	2.80	0.53	60.6	1.45	0.53	25.2	0.60	0.60	0.0	0.0	258,519	20,731	53,118
	0341107	No	3.96	0.83	44.7	5.66	1.63	47.6	0.60	--	0.0	0.0	273,941	21,967	56,287
	1312701	No	0.51	0.07	62.0	2.40	0.80	36.3	0.60	0.00	0.0	0.0	227,108	18,212	46,664
	1722206	No	42.9	15.5	373.	8.63	2.48	39.3	--	--	0.0	--	108,151	8,673	22,222
	2230100	No	5.17	1.24	32.2	1.21	0.46	42.8	--	0.60	--	0.0	213,598	17,128	43,888
	2611101	No	0.73	0.10	52.9	5.22	1.62	75.1	0.30	0.30	0.0	0.0	181,223	14,532	37,236
	2731503	No	8.12	1.99	137.	113.	21.9	5.40	0.20	0.20	0.0	0.0	181,223	14,532	37,236
	3040706	No	2.43	0.43	186.	17.4 *	3.73*	43.5	--	0.60	--	0.0	227,108	18,212	46,664
	0120105	Yes	11.0	2.99	116.	53.8	11.6	34.6	1.6	3.7	0.0	0.0	273,941	21,967	56,287
	0131102	Yes	72.0	36.8	813.	42.5	9.59	60.4	1.5	1.8	0.0	0.0	273,941	21,967	56,287
	0131201	Yes	8.88	2.37	144.	80.9	16.4	109.	0.60	1.9	0.0	10.3	108,151	8,673	22,222
	0251900	Yes	5.40	1.23	269.	13.6	3.47	198.	0.60	0.60	0.0	0.0	273,941	21,967	56,287
	0310201	Yes	4.50	1.12	120.	8.46	2.44	214.	0.60	0.60	0.0	0.0	273,941	21,967	56,287
	0320101	Yes	7.47	1.81	333.	3.57	1.08	209.	1.0	3.1	0.0	0.0	273,941	21,967	56,287
	0321307	Yes	4.01	0.97	161.	0.01	0.01	146.	--	8.4	0.0	0.0	273,941	21,967	56,287
	0351205	Yes	28.9	9.93	706.	1400.	173.	81.4	3.2	3.3	0.0	33.7	258,519	20,731	53,118
	0410100	Yes	13.2	3.62	18.6	16100.	1330.	43.2	2.9	1.4	0.0	0.0	108,151	8,673	22,222
	0411306	Yes	171.	94.4	1240.	6540.	618.	122.	0.50	7.8	0.0	0.0	213,598	17,128	43,888
	0411603	Yes	7.99	2.10	215.	41.0	9.31	115.	7.0	0.60	0.0	0.0	213,598	17,128	43,888
	0520809	Yes	53.5	24.5	543.	3390.	353.	347.	0.40	--	0.0	0.0	108,151	8,673	22,222
	0531301	Yes	33.1	11.8	241.	21.7	5.43	160.	0.40	0.60	0.0	0.0	181,223	14,532	37,236
	0612002	Yes	6.29	1.62	705.	12.4	3.13	135.	0.50	10.	0.0	0.0	213,598	17,128	43,888
	0651901	Yes	1.25	0.19	78.0	105.	19.7	70.9	0.90	0.60	0.0	0.0	108,151	8,673	22,222
	0710103	Yes	5.64	1.24	232.	309. *	34.5 *	217.	0.60	0.60	0.0	0.0	227,108	18,212	46,664
	0750406	Yes	51.6	23.2	667.	11.3	3.02	52.4	0.70	1.4	0.0	0.0	227,108	18,212	46,664
	0821009	Yes	2.25	0.40	97.5	31.5	7.44	90.5	1.1	--	0.0	0.0	227,108	18,212	46,664
	0911503	Yes	37.6	13.7	259.	37.3	7.85	21.7	0.40	0.30	0.0	0.0	227,108	18,212	46,664
	0920801	Yes	1.90	0.34	80.1	0.07	0.04	9.26	0.50	0.30	0.0	0.0	227,108	18,212	46,664
	0921304	Yes	2.21	0.39	131.	309. *	34.5 *	75.8	0.50	0.00	0.0	0.0	227,108	18,212	46,664
	1010503	Yes	136.	71.0	1560.	113.	21.0	7030.	11.	30.	6.3	5.1	213,598	17,128	43,888
	1030204	Yes	3.00	0.57	101.	309. *	34.5 *	65.7	0.50	0.50	0.0	0.0	213,598	17,128	43,888
	1051200	Yes	1.37	0.21	112.	5.65	1.57	142.	0.30	0.40	0.0	0.0	213,598	17,128	43,888
	1120401	Yes	39.6	15.0	248.	8.52	2.28	99.0	0.90	2.6	0.0	6.5	181,223	14,532	37,236
	1121300	Yes	2.15	0.39	60.0	1.46	0.55	144.	0.70	17.	0.0	0.0	213,598	17,128	43,888
	1130806	Yes	10.0	2.75	258.	1.82	0.63	81.0	7.3	8.3	0.0	0.0	213,598	17,128	43,888
	1140508	Yes	15.5	5.58	775.	26.2	5.77	90.0	1.1	4.8	0.0	0.0	181,223	14,532	37,236
	1332402	Yes	8.32	2.26	275.	6.75	1.98	182.	0.60	0.60	0.0	0.0	291,118	23,345	59,816
	1333806	Yes	22.4	7.57	318.	7.76	2.14	61.1	0.60	2.2	0.0	0.0	227,108	18,212	46,664
	1352806	Yes	3.68	0.73	94.2	2.83	0.88	71.3	1.9	0.50	0.0	0.0	227,108	18,212	46,664
	1410406	Yes	4.67	0.98	166.	6.96	1.53	130.	0.60	1.9	0.0	0.0	273,941	21,967	56,287
	1440205	Yes	16.7	5.44	236.	54.7	11.9	24.9	1.4	2.2	0.0	0.0	273,941	21,967	56,287
	1450907	Yes	29.8	11.0	73.5	0.23	0.12	26.0	0.60	0.50	0.0	0.0	258,519	20,731	53,118

Table C1-7. Estimated Environmental Lead Levels in the 1997 Housing Stock, As Determined from National Survey Units. (Continued)

Year Built	National Survey ID	LBP Present?	Wipe Floor Dust-Lead Loading (ug/ft2)	Vac. Floor Dust-Lead Loading (ug/ft2)	BN Floor Dust-Lead Conc. (ug/g)	Wipe W. Sill Dust-Lead Loading (ug/ft2)	Vac. W. Sill Dust-Lead Loading (ug/ft2)	Yardwide Avg. Soil-Lead Conc. (ug/g)	Obs. Max. Interior XRF (mg/cm2)	Obs. Max. Exterior XRF (mg/cm2)	Damaged Interior LBP (ft2)	Damaged Exterior LBP (ft2)	1997 Weight	# Children 12-35 mo.	# Children 12-71 mo.
1940-1959 (cont.)	1521400	Yes	25.0	8.44	173.	130.	24.5	145.	2.4	2.8	6.3	0.0	227,108	18,212	46,664
	1521509	Yes	22.5	7.13	394.	24.2	5.95	132.	1.5	13.	0.0	278.5	227,108	18,212	46,664
	1530500	Yes	3.51	0.69	160.	256.	39.7	264.	1.8	3.7	0.0	56.0	227,108	18,212	46,664
	1550102	Yes	10.0	2.78	287.	2.47	0.81	209.	3.5	2.1	12.5	3.0	227,108	18,212	46,664
	1550607	Yes	17.6	5.29	419.	309. *	34.5 *	145.	1.2	2.0	73.5	0.0	227,108	18,212	46,664
	1551704	Yes	27.8	9.63	314.	58.9	11.7	136.	2.9	2.6	0.0	0.0	227,108	18,212	46,664
	1730407	Yes	12.2	3.62	162.	299.	50.1	63.9	1.8	2.3	4.8	0.0	108,151	8,673	22,222
	1730704	Yes	5.18	1.18	210.	7.43	2.19	77.3	2.1	1.5	0.0	0.0	108,151	8,673	22,222
	1730803	Yes	6.40	1.50	88.4	62.3	13.3	77.3	1.8	1.5	0.0	0.0	108,151	8,673	22,222
	1731603	Yes	0.63	0.09	17.4	309. *	34.5 *	171.	1.5	1.4	0.0	0.0	108,151	8,673	22,222
	1750108	Yes	5.32	1.32	316.	6.47	1.74	53.8	1.2	1.8	0.0	0.0	433,850	34,790	89,143
	1831106	Yes	26.4	9.79	836.	173.	31.5	1410.	2.0	--	0.0	--	111,336	8,928	22,876
	1831304	Yes	14.8	4.31	444.	475.	74.0	1410.	0.60	--	0.0	--	108,151	8,673	22,222
	1840305	Yes	13.8	3.99	244.	177.	31.4	313.	20.	--	0.0	--	108,151	8,673	22,222
	1841105	Yes	17.1	5.37	284.	188.	33.8	313.	1.0	--	0.0	--	111,336	8,928	22,876
	2022705	Yes	13.5	3.73	94.4	15.3	3.98	60.1	0.70	1.5	0.0	0.0	433,850	34,790	89,143
	2030302	Yes	4.07	0.89	102.	5.45	1.68	33.7	0.60	0.60	0.0	0.0	433,850	34,790	89,143
	2110906	Yes	158.	90.2	1680.	475.	71.3	372.	0.60	6.3	0.0	7.3	108,151	8,673	22,222
	2141505	Yes	0.62	0.08	32.0	0.13	0.07	58.9	1.7	1.5	1.4	2.5	291,118	23,345	59,816
	2142107	Yes	5.79	1.35	93.6	309. *	34.5 *	123.	1.2	--	0.0	--	227,108	18,212	46,664
	2211902	Yes	17.9	5.58	61.7	9.66	2.73	22.0	0.70	0.90	0.0	0.0	213,598	17,128	43,888
	2332005	Yes	27.4	9.06	761.	59.9	12.0	313.	8.0	5.0	0.0	77.1	433,850	34,790	89,143
	2343606	Yes	4.20	0.90	136.	1.73	0.64	225.	0.80	2.5	0.0	0.0	433,850	34,790	89,143
	2421709	Yes	7.81	2.14	169.	107.	20.4	52.4	0.60	1.4	0.0	0.0	433,850	34,790	89,143
	2441509	Yes	36.6	13.7	1690.	50.7	11.1	4320.	0.60	3.9	0.0	118.3	411,982	33,037	84,650
	2451805	Yes	3.14	0.63	193.	335.	54.1	34.1	0.60	0.60	0.0	0.0	433,850	34,790	89,143
	2520906	Yes	40.3	14.6	321.	45.3	9.90	55.8	0.80	0.70	0.0	0.0	213,598	17,128	43,888
	2540102	Yes	78.8	34.0	254.	234.	40.6	102.	2.7	1.5	201.9	20.0	213,598	17,128	43,888
	2540201	Yes	4.97	1.09	266.	27.0	6.34	33.0	0.60	1.2	0.0	0.0	213,598	17,128	43,888
	2541407	Yes	56.2	26.5	378.	19.1	4.65	485.	0.60	0.50	0.0	0.0	181,223	14,532	37,236
	2541902	Yes	1.25	0.19	61.0	98.4	19.5	116.	0.70	0.50	0.0	0.0	213,598	17,128	43,888
	2610103	Yes	7.48	1.85	283.	16.9	4.38	43.5	0.20	0.20	0.0	0.0	213,598	17,128	43,888
	2651206	Yes	4.48	1.00	16.8	39.7	8.88	26.3	0.60	0.30	0.0	0.0	213,598	17,128	43,888
	2652303	Yes	4.38	0.98	273.	309. *	34.5 *	49.0	0.50	0.30	0.0	0.0	213,598	17,128	43,888
	2711505	Yes	18.6	5.73	210.	642.	95.8	218.	1.7	7.6	0.0	0.0	181,223	14,532	37,236
	2730703	Yes	2.13	0.39	84.7	9.28	2.64	119.	0.40	0.20	0.0	0.0	213,598	17,128	43,888
	2731800	Yes	19.9	6.70	114.	258.	44.2	12.1	0.40	1.0	0.0	0.0	213,598	17,128	43,888
	2812204	Yes	10.2	3.08	483.	15.1	3.96	162.	2.8	0.60	0.0	0.0	213,598	17,128	43,888
	2840403	Yes	15.2	5.31	1070.	9.29	2.60	52.1	6.1	8.7	0.0	0.0	213,598	17,128	43,888
	2841203	Yes	37.3	15.0	1270.	1290.	159.	61.9	9.6	13.	0.0	0.0	213,598	17,128	43,888
	2841500	Yes	4.84	1.16	118.	0.66	0.28	41.4	1.0	1.8	0.0	0.0	213,598	17,128	43,888
	2910107	Yes	10.1	2.68	230.	3.53	1.15	51.8	1.4	0.50	0.0	0.0	273,941	21,967	56,287
	2931202	Yes	16.9	5.21	316.	40.0	9.10	220.	0.80	0.50	0.0	0.0	108,151	8,673	22,222
	2940708	Yes	4.68	0.98	218.	6.77	2.00	44.3	1.7	1.5	0.0	0.0	273,941	21,967	56,287
	3011509	Yes	4.35	0.89	330.	11.8	3.24	346.	0.60	1.4	0.0	0.0	227,108	18,212	46,664
													19,717,970	1,581,184	4,051,451
1960-1979	0130708	No	3.35	0.83	87.9	32.7	7.31	29.7	0.60	0.60	0.0	0.0	658,726	52,823	135,348
	0131003	No	6.35	2.01	111.	7.53	2.21	5.35	0.60	0.00	0.0	0.0	291,351	23,363	59,864
	0150201	No	12.2	5.99	68.8	11.7	3.22	6.16	0.60	0.50	0.0	0.0	291,351	23,363	59,864
	0330308	No	1.97	0.47	54.8	1.68	0.62	61.6	0.60	0.60	0.0	0.0	291,351	23,363	59,864
	0350306	No	2.65	0.57	112.	4.35	1.31	14.2	--	--	0.0	0.0	658,726	52,823	135,348
	0420901	No	9.30	3.54	20.2	1590.	206.	21.0	0.40	0.60	0.0	0.0	291,351	23,363	59,864
	0430108	No	3.89	1.08	68.8	12.7	3.44	21.3	0.30	0.60	0.0	0.0	116,364	9,331	23,909
	0440305	No	12.1	5.43	245.	3.11	1.03	97.4	0.50	0.60	0.0	0.0	116,364	9,331	23,909
	0440602	No	5.52	1.72	144.	8.69	2.85	79.3	--	--	0.0	0.0	316,764	25,401	65,085
	0541201	No	1.72	0.33	21.5	6.48	1.51	17.9	0.60	0.60	0.0	0.0	116,364	9,331	23,909
	0940700	No	1.79	0.34	47.0	81.5 *	12.2 *	7.23	0.30	0.30	0.0	0.0	291,351	23,363	59,864
	0940809	No	6.32	1.93	429.	1.00	0.40	17.7	0.30	0.30	0.0	0.0	291,351	23,363	59,864
	1020205	No	2.30	0.51	171.	15.0	3.97	49.2	0.30	0.50	0.0	0.0	316,764	25,401	65,085

Table C1-7. Estimated Environmental Lead Levels in the 1997 Housing Stock, As Determined from National Survey Units. (Continued)

Year Built	National Survey ID	LBP Present?	Wipe Floor Dust-Lead Loading (ug/ft2)	Vac. Floor Dust-Lead Loading (ug/ft2)	BN Floor Dust-Lead Conc. (ug/g)	Wipe W. Sill Dust-Lead Loading (ug/ft2)	Vac. W. Sill Dust-Lead Loading (ug/ft2)	Yardwide Avg. Soil-Lead Conc. (ug/g)	Obs. Max. Interior XRF (mg/cm2)	Obs. Max. Exterior XRF (mg/cm2)	Damaged Interior LBP (ft2)	Damaged Exterior LBP (ft2)	1997 Weight	# Children 12-35 mo.	# Children 12-71 mo.
1960-1979 (cont.)	1020502	No	3.30	0.78	160.	19.5	4.52	58.3	0.30	--	0.0	0.0	316,764	25,401	65,085
	1021005	No	1.74	0.33	198.	4.60	1.46	25.5	0.50	0.30	0.0	0.0	316,764	25,401	65,085
	1040500	No	3.46	0.91	208.	9.64	2.73	24.5	0.40	--	0.0	0.0	116,364	9,331	23,909
	1323609	No	1.37	0.27	102.	81.5 *	12.2 *	20.4	0.60	--	0.0	--	312,998	25,099	64,312
	1441302	No	1.06	0.18	85.2	0.02	0.01	13.0	0.60	0.50	0.0	0.0	658,726	52,823	135,348
	2220507	No	5.81	1.85	123.	81.5 *	12.2 *	14.1	0.60	0.60	0.0	0.0	316,764	25,401	65,085
	2230209	No	2.00	0.39	68.6	4.60	1.38	5.58	0.60	0.60	0.0	0.0	316,764	25,401	65,085
	2511806	No	1.85	0.37	52.2	0.83	0.34	11.6	0.60	0.50	0.0	0.0	116,364	9,331	23,909
	2521201	No	12.9	5.58	183.	127.	24.2	73.4	0.60	0.10	0.0	0.0	316,764	25,401	65,085
	2551000	No	1.29	0.22	52.1	2.05	0.73	22.6	0.60	0.50	0.0	0.0	316,764	25,401	65,085
	2552107	No	1.47	0.25	40.5	0.52	0.23	27.2	0.60	0.50	0.0	0.0	316,764	25,401	65,085
	2822005	No	2.68	0.58	65.8	124.	23.7	82.5	0.60	0.00	0.0	0.0	316,764	25,401	65,085
	2831006	No	1.21	0.19	33.8	0.12	0.07	21.1	0.60	0.60	0.0	0.0	316,764	25,401	65,085
	2831709	No	3.01	0.73	64.3	6.10	1.85	40.8	0.60	0.60	0.0	0.0	116,364	9,331	23,909
	3050101	No	3.81	0.98	458.	81.5 *	12.2 *	6.68	0.60	0.60	0.0	0.0	312,998	25,099	64,312
	0130906	Yes	5.38	1.56	68.2	5.40	1.67	39.5	0.60	0.60	0.0	0.0	126,372	10,134	25,966
	0150102	Yes	7.59	2.72	188.	2.64	0.91	4.79	0.80	0.60	0.0	0.0	658,726	52,823	135,348
	0250902	Yes	3.61	0.97	206.	4.29	1.28	180.	0.60	0.60	0.0	0.0	352,318	28,252	72,391
	0252404	Yes	8.08	3.11	225.	329.	51.9	604.	1.0	0.60	0.0	0.0	291,351	23,363	59,864
	0311209	Yes	5.95	1.80	330.	885.	126.	186.	0.80	0.60	0.0	0.0	352,318	28,252	72,391
	0331009	Yes	3.08	0.78	27.5	8.63	2.48	15.3	0.60	0.60	0.0	0.0	352,318	28,252	72,391
	0340505	Yes	6.94	2.76	15.8	0.51	0.22	23.7	0.60	0.60	0.0	0.0	658,726	52,823	135,348
	0340802	Yes	11.5	6.33	652.	6.86	2.03	31.8	0.90	0.60	0.0	0.0	352,318	28,252	72,391
	0341404	Yes	2.92	0.74	34.1	28.0	6.73	20.0	1.0	0.60	0.0	0.0	658,726	52,823	135,348
	0410605	Yes	6.39	2.07	60.9	605.	91.1	127.	1.4	0.00	0.0	0.0	291,351	23,363	59,864
	0421206	Yes	41.7	31.7	643.	665.	98.7	22.8	0.50	1.7	0.0	0.0	291,351	23,363	59,864
	0430207	Yes	13.2	5.38	87.3	15.8	4.14	35.2	0.50	0.80	0.0	0.0	316,764	25,401	65,085
	0430306	Yes	13.5	6.81	7.04	6.96	2.07	27.2	0.40	0.70	0.0	0.0	316,764	25,401	65,085
	0430702	Yes	12.5	6.50	318.	217. *	28.3 *	26.4	0.40	0.40	0.0	0.0	116,364	9,331	23,909
	0440107	Yes	15.7	7.85	319.	130.	24.1	34.7	0.60	0.50	0.0	0.0	316,764	25,401	65,085
	0441105	Yes	4.40	1.19	75.4	10.7	2.98	5.22	0.50	0.60	0.0	0.0	116,364	9,331	23,909
	0441204	Yes	19.4	9.48	177.	326.	45.9	87.4	0.40	0.60	0.0	0.0	316,764	25,401	65,085
	0530105	Yes	7.27	2.35	246.	38.2	8.68	50.9	1.4	0.70	0.0	0.0	116,364	9,331	23,909
	0530600	Yes	4.18	1.15	193.	217. *	28.3 *	215.	1.0	0.60	0.0	0.0	291,351	23,363	59,864
	0531400	Yes	32.3	20.0	143.	103.	19.9	56.1	0.70	1.7	0.0	0.0	116,364	9,331	23,909
	0540203	Yes	8.59	3.31	59.9	217. *	28.3 *	14.8	0.80	0.70	0.0	0.0	316,764	25,401	65,085
	0541300	Yes	7.59	2.62	184.	160.	26.8	7.52	0.90	0.60	0.0	0.0	316,764	25,401	65,085
	0621607	Yes	2.37	0.49	63.2	0.45	0.20	39.4	0.70	0.30	0.0	0.0	126,372	10,134	25,966
	0631408	Yes	3.46	0.93	221.	26.2	6.36	85.4	0.40	--	0.0	0.0	126,372	10,134	25,966
	0840702	Yes	3.23	0.77	82.4	1.20	0.47	30.6	0.80	1.2	0.0	5.9	291,351	23,363	59,864
	0911404	Yes	1.71	0.30	98.6	1.76	0.61	29.8	0.30	0.30	0.0	0.0	173,719	13,931	35,694
	0930701	Yes	3.58	1.00	378.	1.53	0.57	19.7	0.60	0.30	0.0	0.0	312,998	25,099	64,312
	1011709	Yes	3.19	0.78	366.	0.45	0.20	996.	11.	0.40	28.8	0.0	116,364	9,331	23,909
	1020304	Yes	2.25	0.47	150.	217. *	28.3 *	26.6	0.80	0.70	0.0	0.0	316,764	25,401	65,085
	1020403	Yes	2.34	0.47	104.	0.81	0.34	25.0	0.70	0.30	0.0	0.0	316,764	25,401	65,085
	1020700	Yes	0.75	0.09	51.5	0.19	0.10	23.8	0.60	0.70	0.0	0.0	316,764	25,401	65,085
	1020809	Yes	2.65	0.57	263.	9.64	2.73	25.4	0.40	--	0.0	0.0	316,764	25,401	65,085
	1050509	Yes	1.02	0.15	44.1	217. *	28.3 *	116.	3.0	0.60	0.0	0.0	316,764	25,401	65,085
	1050608	Yes	1.09	0.20	124.	0.12	0.07	57.5	0.30	0.30	0.0	0.0	316,764	25,401	65,085
	1051408	Yes	1.50	0.27	127.	217. *	28.3 *	143.	0.30	1.7	0.0	0.0	116,364	9,331	23,909
	1150200	Yes	3.17	0.97	241.	60.6	12.9	35.3	1.0	9.1	0.0	0.0	291,351	23,363	59,864
	1150705	Yes	3.62	0.92	185.	450.	70.8	81.6	1.6	0.40	0.0	0.0	291,351	23,363	59,864
	1241801	Yes	2.48	0.51	112.	3440.	377.	196.	0.60	1.4	0.0	0.0	451,561	36,211	92,782
	1311505	Yes	1.37	0.25	81.8	1.73	0.62	20.8	0.60	0.00	0.0	0.0	173,719	13,931	35,694
	1312800	Yes	12.3	6.71	250.	0.50	0.20	13.8	0.90	--	0.0	--	312,998	25,099	64,312
	1322601	Yes	1.37	0.22	101.	0.74	0.31	33.3	0.60	--	0.0	--	312,998	25,099	64,312
	1353309	Yes	0.66	0.09	15.8	0.28	0.12	51.6	0.60	0.00	0.0	0.0	291,351	23,363	59,864
	1441005	Yes	1.93	0.36	46.9	0.83	0.34	18.8	1.5	0.50	0.0	0.0	658,726	52,823	135,348
	1510403	Yes	4.40	1.19	249.	217. *	28.3 *	4.63	0.50	0.60	0.0	0.0	173,719	13,931	35,694
	1510908	Yes	7.86	3.15	188.	1.87	0.66	14.8	0.30	0.20	0.0	0.0	312,998	25,099	64,312
	1520204	Yes	4.44	1.50	227.	7.31	2.16	35.9	0.30	10.	0.0	0.0	312,998	25,099	64,312

Table C1-7. Estimated Environmental Lead Levels in the 1997 Housing Stock, As Determined from National Survey Units. (Continued)

Year Built	National Survey ID	LBP Present?	Wipe Floor Dust-Lead Loading (ug/ft2)	Vac. Floor Dust-Lead Loading (ug/ft2)	BN Floor Dust-Lead Conc. (ug/g)	Wipe W. Sill Dust-Lead Loading (ug/ft2)	Vac. W. Sill Dust-Lead Loading (ug/ft2)	Yardwide Avg. Soil-Lead Conc. (ug/g)	Obs. Max. Interior XRF (mg/cm2)	Obs. Max. Exterior XRF (mg/cm2)	Damaged Interior LBP (ft2)	Damaged Exterior LBP (ft2)	1997 Weight	# Children 12-35 mo.	# Children 12-71 mo.
1960-1979 (cont.)	1530104	Yes	7.38	2.40	204.	51.3	11.3	78.7	3.3	0.10	0.0	0.0	312,998	25,099	64,312
	1530302	Yes	6.81	2.11	328.	448.	66.2	68.4	0.90	1.5	0.0	0.0	312,998	25,099	64,312
	1530807	Yes	11.0	4.49	238.	48.3	9.95	40.5	0.60	2.5	0.0	0.0	312,998	25,099	64,312
	1531607	Yes	12.5	5.19	289.	5790.	618.	105.	22.	11.	12.5	27.5	173,719	13,931	35,694
	1531706	Yes	6.26	2.07	159.	545.	75.7	23.4	0.00	1.3	0.0	0.0	312,998	25,099	64,312
	1540202	Yes	2.26	0.49	139.	217. *	28.3 *	15.9	0.00	0.70	0.0	0.0	173,719	13,931	35,694
	1540400	Yes	8.48	3.47	159.	16.6	4.32	49.9	0.60	0.20	0.0	0.0	312,998	25,099	64,312
	1540806	Yes	13.1	5.97	175.	217. *	28.3 *	30.1	0.70	0.30	0.0	0.0	312,998	25,099	64,312
	1541200	Yes	3.28	0.82	180.	217. *	28.3 *	17.1	0.70	0.70	0.0	0.0	173,719	13,931	35,694
	1741701	Yes	40.0	31.9	141.	217. *	28.3 *	54.7	1.0	0.80	0.0	0.0	243,025	19,488	49,934
	1741800	Yes	10.5	4.15	143.	217. *	28.3 *	95.7	2.5	0.90	9.5	0.0	451,561	36,211	92,782
	1743103	Yes	1.21	0.19	29.1	0.07	0.04	28.6	1.0	3.3	0.0	0.0	451,561	36,211	92,782
	2040301	Yes	5.81	1.68	126.	22.8	5.53	14.8	0.60	3.6	0.0	0.0	451,561	36,211	92,782
	2122000	Yes	23.1	14.2	395.	132.	22.9	355.	1.4	0.60	0.0	0.0	291,351	23,363	59,864
	2130706	Yes	9.68	3.49	127.	9.59	2.72	13.7	1.5	1.3	57.3	0.0	173,719	13,931	35,694
	2131902	Yes	2.60	0.57	76.4	50.1	11.0	21.1	0.60	1.6	0.0	0.6	312,998	25,099	64,312
	2141604	Yes	9.93	3.72	87.0	3.70	1.21	39.2	1.1	0.60	0.0	0.0	173,719	13,931	35,694
	2151207	Yes	8.75	3.48	324.	42.3	8.63	17.5	1.2	0.50	0.0	0.0	312,998	25,099	64,312
	2211308	Yes	4.67	1.32	89.4	503.	75.4	20.4	0.70	1.3	0.0	8.0	316,764	25,401	65,085
	2230506	Yes	2.83	0.64	77.3	12.2	3.14	6.11	0.90	--	0.0	0.0	316,764	25,401	65,085
	2351500	Yes	5.26	1.56	135.	305.	45.8	115.	1.2	0.10	1.1	0.0	243,025	19,488	49,934
	2352201	Yes	3.14	0.88	180.	217. *	28.3 *	42.5	1.1	0.10	0.0	0.0	243,025	19,488	49,934
	2430403	Yes	4.76	1.47	457.	4.19	1.22	69.7	0.60	3.4	0.0	72.4	451,561	36,211	92,782
	2431807	Yes	4.69	1.70	215.	6.17	1.86	41.1	0.90	5.1	0.0	0.0	451,561	36,211	92,782
	2452605	Yes	4.79	1.40	315.	217. *	28.3 *	121.	1.0	0.60	0.0	0.0	451,561	36,211	92,782
	2520609	Yes	6.30	2.55	803.	0.22	0.11	15.7	0.60	0.50	0.0	0.0	116,364	9,331	23,909
	2521102	Yes	4.29	1.35	117.	149.	27.7	26.8	0.60	1.0	0.0	0.0	316,764	25,401	65,085
	2531804	Yes	1.31	0.21	60.0	0.42	0.19	66.4	1.2	0.50	0.0	0.0	316,764	25,401	65,085
	2541506	Yes	35.7	26.0	2200.	315.	50.2	45.2	4.6	0.50	12.8	0.0	116,364	9,331	23,909
	2620508	Yes	2.24	0.45	133.	217. *	28.3 *	46.1	0.40	0.60	0.0	0.0	126,372	10,134	25,966
	2621704	Yes	4.53	1.33	191.	1.40	0.53	68.5	0.40	8.8	0.0	0.0	126,372	10,134	25,966
	2622603	Yes	1.28	0.21	2.01	3.01	1.02	54.6	0.50	1.1	0.0	0.0	291,351	23,363	59,864
	2623007	Yes	1.19	0.19	130.	1.76	0.64	26.0	0.60	0.60	0.0	0.0	126,372	10,134	25,966
	2650208	Yes	2.56	0.60	93.0	14.9	3.90	52.7	0.50	3.0	0.0	6.0	116,364	9,331	23,909
	2711109	Yes	7.60	2.68	128.	32.0	7.54	32.0	0.70	0.80	0.0	0.0	316,764	25,401	65,085
	2751402	Yes	106.	124.	50400.	19.9	4.66	35.0	0.30	0.60	0.0	0.0	291,351	23,363	59,864
	2810307	Yes	3.33	0.78	137.	217. *	28.3 *	23.2	0.50	0.00	0.0	0.0	291,351	23,363	59,864
	2812105	Yes	3.39	0.79	137.	217. *	28.3 *	91.3	1.2	1.5	0.0	0.0	291,351	23,363	59,864
	2830602	Yes	5.12	1.72	170.	11.4	3.16	32.1	0.60	1.5	0.0	0.0	116,364	9,331	23,909
	2832004	Yes	6.16	1.95	283.	0.34	0.15	20.8	0.60	0.60	0.0	0.0	116,364	9,331	23,909
	2832103	Yes	1.74	0.31	87.5	217. *	28.3 *	75.6	0.20	0.20	0.0	0.0	316,764	25,401	65,085
	2840106	Yes	7.67*	4.25*	740. *	217. *	28.3 *	63.4	0.80	5.1	0.0	20.7	126,372	10,134	25,966
	2840205	Yes	7.67*	4.25*	740. *	217. *	28.3 *	63.4	0.70	2.0	0.0	25.7	316,764	25,401	65,085
	2841401	Yes	2.57	0.60	59.8	228.	35.9	35.6	0.60	1.6	0.0	0.0	116,364	9,331	23,909
	2940401	Yes	1.38	0.22	66.4	1.31	0.50	27.2	1.2	--	0.0	--	658,726	52,823	135,348
	3051000	Yes	1.95	0.37	152.	217. *	28.3 *	31.1	0.70	0.60	0.0	0.0	312,998	25,099	64,312
													34,984,547	2,805,411	7,188,275
>1979	0130708	No	3.35	0.83	87.9	32.7	7.31	29.7	0.60	0.60	0.0	0.0	889,038	71,292	182,671
	0131003	No	6.35	2.01	111.	7.53	2.21	5.35	0.60	0.00	0.0	0.0	889,038	71,292	182,671
	0150201	No	12.2	5.99	68.8	11.7	3.22	6.16	0.60	0.50	0.0	0.0	889,038	71,292	182,671
	0330308	No	1.97	0.47	54.8	1.68	0.62	61.6	0.60	0.60	0.0	0.0	889,038	71,292	182,671
	0350306	No	2.65	0.57	112.	4.35	1.31	14.2	--	--	0.0	0.0	889,038	71,292	182,671
	0420901	No	9.30	3.54	20.2	1590.	206.	21.0	0.40	0.60	0.0	0.0	889,038	71,292	182,671
	0430108	No	3.89	1.08	68.8	12.7	3.44	21.3	0.30	0.60	0.0	0.0	889,038	71,292	182,671
	0440305	No	12.1	5.43	245.	3.11	1.03	97.4	0.50	0.60	0.0	0.0	889,038	71,292	182,671
	0440602	No	5.52	1.72	144.	8.69	2.85	79.3	--	--	0.0	0.0	889,038	71,292	182,671
	0541201	No	1.72	0.33	21.5	6.48	1.51	17.9	0.60	0.60	0.0	0.0	889,038	71,292	182,671
	0940700	No	1.79	0.34	47.0	83.0 *	12.3 *	7.23	0.30	0.30	0.0	0.0	889,038	71,292	182,671
	0940809	No	6.32	1.93	429.	1.00	0.40	17.7	0.30	0.30	0.0	0.0	889,038	71,292	182,671

Table C1-7. Estimated Environmental Lead Levels in the 1997 Housing Stock, As Determined from National Survey Units. (Continued)

Year Built	National Survey ID	LBP Present?	Wipe Floor Dust-Lead Loading (ug/ft2)	Vac. Floor Dust-Lead Loading (ug/ft2)	BN Floor Dust-Lead Conc. (ug/g)	Wipe W. Sill Dust-Lead Loading (ug/ft2)	Vac. W. Sill Dust-Lead Loading (ug/ft2)	Yardwide Avg. Soil-Lead Conc. (ug/g)	Obs. Max. Interior XRF (mg/cm2)	Obs. Max. Exterior XRF (mg/cm2)	Damaged Interior LBP (ft2)	Damaged Exterior LBP (ft2)	1997 Weight	# Children 12-35 mo.	# Children 12-71 mo.
>1979	1020205	No	2.30	0.51	171.	15.0	3.97	49.2	0.30	0.50	0.0	0.0	889,038	71,292	182,671
(cont.)	1020502	No	3.30	0.78	160.	19.5	4.52	58.3	0.30	--	0.0	0.0	889,038	71,292	182,671
	1021005	No	1.74	0.33	198.	4.60	1.46	25.5	0.50	0.30	0.0	0.0	889,038	71,292	182,671
	1040500	No	3.46	0.91	208.	9.64	2.73	24.5	0.40	--	0.0	0.0	889,038	71,292	182,671
	1323609	No	1.37	0.27	102.	83.0 *	12.3 *	20.4	0.60	--	0.0	--	889,038	71,292	182,671
	1441302	No	1.06	0.18	85.2	0.02	0.01	13.0	0.60	0.50	0.0	0.0	889,038	71,292	182,671
	2220507	No	5.81	1.85	123.	83.0 *	12.3 *	14.1	0.60	0.60	0.0	0.0	889,038	71,292	182,671
	2230209	No	2.00	0.39	68.6	4.60	1.38	5.58	0.60	0.60	0.0	0.0	889,038	71,292	182,671
	2511806	No	1.85	0.37	52.2	0.83	0.34	11.6	0.60	0.50	0.0	0.0	889,038	71,292	182,671
	2521201	No	12.9	5.58	183.	127.	24.2	73.4	0.60	0.10	0.0	0.0	889,038	71,292	182,671
	2551000	No	1.29	0.22	52.1	2.05	0.73	22.6	0.60	0.50	0.0	0.0	889,038	71,292	182,671
	2552107	No	1.47	0.25	40.5	0.52	0.23	27.2	0.60	0.50	0.0	0.0	889,038	71,292	182,671
	2822005	No	2.68	0.58	65.8	124.	23.7	82.5	0.60	0.00	0.0	0.0	889,038	71,292	182,671
	2831006	No	1.21	0.19	33.8	0.12	0.07	21.1	0.60	0.60	0.0	0.0	889,038	71,292	182,671
	2831709	No	3.01	0.73	64.3	6.10	1.85	40.8	0.60	0.60	0.0	0.0	889,038	71,292	182,671
	3050101	No	3.81	0.98	458.	83.0 *	12.3 *	6.68	0.60	0.60	0.0	0.0	889,038	71,292	182,671
													24,893,064	1,996,175	5,114,778
												TOTAL ACROSS ALL UNITS:	99,271,901	7,960,614	20,397,397

* As no data for this parameter existed in the National Survey database for the given housing unit, this value is the average of the values across all units in the same year-built category and having the same value for the LBP indicator that had reported data (see Table 3-14 in Chapter 3). The average is weighted using the 1997 weights.

Note: Dust-lead loadings are area-weighted arithmetic averages for the unit. "Wipe" loadings are converted from Blue Nozzle ("Vac.") vacuum loadings (see Chapter 4). Dust-lead concentrations are mass-weighted arithmetic averages of individual sample concentrations for the unit that have been adjusted for low tap weights (USEPA, 1996c). Soil-lead concentration represents a weighted arithmetic yardwide average for the unit, with remote sample results weighted twice that of entryway and dripline samples.

APPENDIX C2

Method for Computing Confidence Intervals Associated with Estimates in the Exposure Assessment and Risk Characterization

APPENDIX C2

METHOD FOR COMPUTING CONFIDENCE INTERVALS ASSOCIATED WITH ESTIMATES IN THE EXPOSURE ASSESSMENT AND RISK CHARACTERIZATION

In Chapters 3 and 5, approximate 95% confidence intervals were calculated for selected exposure and risk estimates to provide a measure of precision for these estimates. These risk estimates included children's geometric mean blood-lead concentrations in the nation's housing stock (Tables 3-36, 3-38, 3-39, and 3-40), the percentage of children's blood-lead concentrations greater than or equal to specified thresholds (10 or 20 µg/dL) (Tables 3-37, 3-38, 3-39, 3-40, 5-1, and 5-9), the percentage of children experiencing IQ decrements greater than or equal to 1, 2, or 3, as a result of lead exposure (Tables 5-1 and 5-9), and average IQ decrement due to childhood lead exposure. Confidence intervals were also computed for lead levels in dust which, when assuming fixed lead levels in other media, control the percentage of children with blood-lead concentrations greater than or equal to 10 µg/dL to specified levels (Tables 5-6 and 5-7). This appendix presents the methodology used to compute these intervals.

For endpoints estimated from the NHANES III data, the method for computing a confidence interval needs to account for the complex survey design associated with NHANES III. To do this, Software for Survey Data Analysis (SUDAAN) was used to compute standard errors for NHANES III distribution parameters. These standard errors were then used to compute standard errors for the estimated exposure and risk endpoints. These methods are presented in the following subsections, according to the type of baseline risk estimate.

1.0 GEOMETRIC MEAN BLOOD-LEAD CONCENTRATION

Data from Phase 2 of NHANES III were used to construct estimates of geometric mean blood-lead concentration for specified subgroups of the nation's children (e.g., 1-2 year old children, 1-2 year old children living in pre-1946 housing). For some confidence level α (from 0 to 1), a $(1-\alpha)*100\%$ confidence interval for the geometric mean was calculated as

$$\left(e^{\ln(GM)-t_{(n-1,\alpha)}s},\ e^{\ln(GM)+t_{(n-1,\alpha)}s}\right) \qquad \textbf{(1)}$$

where GM is the estimated geometric mean blood-lead concentration of the subgroup of interest, s is the standard error of the arithmetic mean of the log-transformed blood-lead concentrations in this subgroup (computed using SUDAAN), n is the sample size for the subgroup, and $t_{(n-1,\alpha)}$ is the $(1-\alpha)*100$th percentile of the Student t-distribution with n-1 degrees of freedom. In the risk analysis, $\alpha=0.05$ (i.e., 95% confidence intervals were calculated). Note that this approach to calculating a confidence interval assumes that blood-lead concentrations are lognormally distributed. While the estimate of s accounted for the complex survey design, the degrees of

freedom for the t-statistic were not adjusted. However, not adjusting the degrees of freedom is anticipated to have little effect because of the large sample size (987).

2.0 PERCENTAGE OF CHILDREN'S BLOOD-LEAD CONCENTRATIONS GREATER THAN OR EQUAL TO A SPECIFIED THRESHOLD

Data from Phase 2 of NHANES III were used to compute estimates of the percentage of children's blood-lead concentrations greater than or equal to a specified threshold for specified subgroups of the nation's children. In the risk analysis, two methods for characterizing the distribution of blood-lead concentrations were used:

Method 1: The distribution was characterized empirically from the observed NHANES III data. Under this method, which produced the estimates presented in Section 3.4.1 of Chapter 3, the estimated percentage equaled the observed percentage of children in the survey who were at or above the threshold, with each child weighted by his/her assigned sampling weight.

Method 2: The percentages were computed using the geometric mean and geometric standard deviation estimated in Method 1 assuming that the distribution of blood-lead concentrations is lognormal. This method was used to compute the estimates presented in Section 5.1.1 of Chapter 5.

Standard errors of the percentages estimated under Method 1 were calculated using SUDAAN to account for the complex survey design in NHANES III. If p_x is the estimated percentage of children with blood-lead concentration at or above X µg/dL (for some threshold X) and $SE(p_x)$ is the estimated standard error of this percentage, then (asymmetric) approximate (1-α)*100% confidence intervals associated with p_x were calculated as

$$\left(e^{\ln(p_x)-t_{(n-1,\alpha)}(SE(p_x)/p_x)},\ e^{\ln(p_x)+t_{(n-1,\alpha)}(SE(p_x)/p_x)}\right) \qquad \textbf{(2)}$$

where $t_{(n-1,\alpha)}$ is the (1-α)*100th percentile of the Student t-distribution with n degrees of freedom (Kleinbaum et.al., 1982).

Under Method 2, the value of p_x is estimated as

$$p_x = 100 - 100 * \left(\frac{\ln(X)-LGM}{\sqrt{LGV}}\right) \qquad \textbf{(3)}$$

where LGM and LGV are the weighted arithmetic mean and variance, respectively, of the log-transformed blood-lead concentrations. Assuming independence between LGM and LGV, and using the first order Taylor series approximation of equation (3), an estimate of variability associated with p_x is

$$var(p_x) = \frac{100^2}{LGV} \; ^2\left(\frac{\ln(X) - LGM}{\sqrt{LGV}} \right) [var(LGM) + \left(\frac{\ln(X) - LGM}{2 * LGV} \right)^2 var(LGV)] \quad \textbf{(4)}$$

The variance of LGM, var(LGM), was estimated in SUDAAN to account for the complex survey design employed in NHANES III. The variability associated with LGV, var(LGV) was estimated based on the chi-squared distribution and the "design effect" for LGV (DE_{LGV}):

$$var(LGV) = \frac{2 * LGV^2}{n-1} * DE_{LGV} \quad \textbf{(5)}$$

where n is the sample size for the subgroup of interest. The design effect for a given statistic quantifies the information lost due to the survey design employed and is calculated as the variance of the statistic assuming the complex survey design was employed in data collection, divided by the variance assuming simple random sampling was employed. Because a design effect for LGV was not easily available, the design effect for LGM was used. Although not the optimal solution, this was deemed more appropriate than not accounting for the complex survey design at all.

Because many of the percentage estimates were small, asymmetric confidence intervals were also calculated for model-based estimates using the logarithmic transformation, and the t-distribution with n-1 degrees of freedom (see Equation 2).

3.0 PERCENTAGE OF CHILDREN WITH IQ DECREMENTS GREATER THAN OR EQUAL TO 1, 2, OR 3

Data from Phase 2 of NHANES III were combined with an estimate of the relationship between blood-lead concentration and IQ decrements based on Schwartz, 1994 (Section 4.4 and Appendix D2) to construct estimates of percentage of children with IQ decrements greater than or equal to 1, 2, or 3 that results from lead exposure. Using notation from Section 2.0 above, estimates of this population characteristic were constructed assuming that blood-lead concentrations are lognormally distributed and that the relationship between blood-lead concentration and IQ score decrements is linear:

$$Percent[IQ\ \ Decrement \geq X] \; = \; p_{X/m} \; = \; 100 \; - \; 100 * \; \left(\frac{\ln(X/m) - \ln(GM)}{\ln(GSD)} \right) \quad \textbf{(6)}$$

where X is the specified IQ decrement, m is the slope of the assumed linear relationship between blood-lead concentration and IQ score decrements, GM is the geometric mean blood-lead concentration for the subgroup, and GSD is the geometric standard deviation of blood-lead concentrations.

The standard error of the percentage in equation (6), necessary for calculating a confidence interval for this percentage, was calculated using a first order Taylor series approximation, using estimates of the variability associated with the values of GM, GSD, and the slope factor m. Thus, the confidence interval considers sampling variability from NHANES III data, as well as variability associated with the blood-lead concentration-IQ decrement relationship.

The function for computing the percentage of children with IQ decrements greater than 1, 2, or 3 was expanded in an alternative parameterization to simplify the procedure:

$$Percent[IQ\ \ Decrements \geq X] = 100 - 100 * \left(\frac{\ln(X) - \ln(m) - LGM}{\sqrt{(LGV)}} \right) \quad \textbf{(7)}$$

where LGM and LGV are the arithmetic mean and variance of the log-transformed blood-lead concentrations. Assuming independence between LGM, LGV, and m and using the first order Taylor series approximation to equation (7), the variability associated with estimated percentage of children with IQ decrements greater than X is

$$var(Percent[IQ\ \ Decrements \geq X]) =$$
$$\frac{100^2}{LGV}\ ^2\left(\frac{\ln(X/m) - LGM}{\sqrt{LGV}} \right)[\frac{1}{m^2} var(m) + var(LGM) + \left(\frac{\ln(X/m) - LGM}{2LGV} \right)^2 var(LGV)] \quad \textbf{(8)}$$

The variance of LGM was estimated in SUDAAN to account for the complex survey design employed in NHANES III. The variability associated with LGV was estimated as described in Section 2.0 of this appendix. The variability associated with m was assumed to be 0.041, based on the meta-analysis described in Schwartz, 1994 (Appendix D2).

Because many of the percentage estimates were small, asymmetric confidence intervals were calculated using the logarithmic transformation (Kleinbaum, et al., 1982) as described in Section 2.0.

4.0 AVERAGE IQ DECREMENT

Data from Phase 2 of NHANES III were combined with an estimate of the relationship between blood-lead concentration and IQ decrements based on Schwartz, 1994 (Section 4.4 and Appendix D2) to construct estimates of average IQ decrement. Estimates of this population characteristic were constructed assuming that blood-lead concentrations are lognormally distributed and that the relationship between blood-lead concentration and IQ score decrements is linear:

$$Average\ IQ\ Decrement\ = m * e^{(LGM + LGV/2)} \quad \textbf{(9)}$$

where m is the slope of the assumed linear relationship between blood-lead concentration and IQ score decrements and LGM and LGV are the weighted arithmetic mean and variance of the log-transformed blood-lead concentrations, respectively.

The standard error of average IQ decrement, necessary to calculate a confidence interval, was calculated using a first order Taylor series approximation and estimates of the variability associated with the values of GM, GSD, and the slope factor m. Thus, confidence intervals presented include sampling variability from NHANES III data, as well as variability associated with the blood-lead concentration-IQ decrement relationship.

Assuming independence between LGM, LGV, and m and using the first order Taylor series approximation to equation (9), the variability associated with estimated average IQ decrement is

$$var(Average\ \ IQ\ \ Decrement) = e^{2*LGM+LGV}(var(m) + m^2 * var(LGM) + \frac{m^2}{4} * var(LGV)) \qquad \textbf{(10)}$$

The variance of LGM was estimated in SUDAAN to account for the complex survey design employed in NHANES III. The variability associated with LGV was estimated as described in Section 2.0 of this appendix. The variability associated with m was assumed to be 0.041, based on the meta-analysis described in Schwartz 1994 (Appendix D2).

Confidence intervals were constructed using the t-distribution with degrees of freedom approximated by one less than the sample size.

5.0 INDIVIDUAL RISKS

Upper confidence bounds on the dust-lead loading which, assuming fixed lead levels in other media, controls the percentage of children with blood-lead concentrations greater than or equal to 10 µg/dL due to exposure at these levels, were calculated and presented in Section 5.3 (Tables 5-6 and 5-7) of the risk analysis. The method used to calculate these upper confidence bounds accounts for the variability associated with estimating the parameters of the Rochester multimedia model, which was used to estimate the dust-lead loading.

The method is presented for the example of predicting the floor dust-lead loading which, assuming fixed lead levels in soil and window sill dust, controls the percentage of children with blood-lead concentrations greater than or equal to 10 µg/dL to no higher than α%. This floor dust-lead loading was estimated as

$$PbF = e^{[\ln(10) - \ln(1.6) * \ ^{-1}(1-\alpha/100) - \beta_0 - \ln(PbS)*\beta_{soil} - \ln(PbWS)*\beta_{windowsill}]/\beta_{floor}} \qquad \textbf{(11)}$$

where $^{-1}$ is the inverse normal transformation, PbF is the floor dust-lead loading, PbS is the soil-lead concentration, PbWS is the window sill dust-lead loading and the β's are estimates of the

coefficients for the Rochester multimedia model (Section 4.2.3). The variance of PbF was calculated using a first order Taylor Series approximation, considering the covariance between the parameter estimates from the Rochester multimedia model:

$$var(PbF) = PbF^2[\frac{1}{z^2}var(y) - 2 * \frac{y}{z^3} * cov(y,z) + \frac{y^2}{z^4} * var(z)] \quad \textbf{(12)}$$

where

$$y = \beta_0 + \ln(PbS) * \beta_{soil} + \ln(PbWS) * \beta_{windowsill} \quad \textbf{(13)}$$

$z = \beta_{floor}$, PbS is the soil-lead concentration, and PbWS is the window sill dust-lead loading. Approximate 95% upper confidence bounds for PbF were then computed as

$$PbF + 1.65 * \sqrt{var(PbF)} \quad \textbf{(14)}$$

In the same manner, this approach was used to calculate upper confidence bounds for the window sill dust-lead loading which controls the percentage of children's blood-lead concentrations greater than or equal to 10 µg/dL, assuming fixed soil-lead concentrations and floor dust-lead loadings.

APPENDIX D1

Assumptions and Scientific Evidence to Account for the Effect of Pica for Paint

APPENDIX D1

ASSUMPTIONS AND SCIENTIFIC EVIDENCE TO ACCOUNT FOR THE EFFECT OF PICA FOR PAINT

The scientific evidence on paint chip ingestion is scant and can be contradictory. It is well known that pica for paint and plaster is associated with lead poisoning. However, survey data and blood-lead concentrations collected in the Rochester Lead-in-Dust Study (USHUD, 1995a) indicated that children whose parents responded that they have a tendency to eat paint chips had blood-lead levels only slightly more elevated, on average, than those who do not exhibit pica. The scientific evidence and assumptions required to estimate the percentage of children who exhibit pica for paint and their blood-lead levels are summarized in this section.

PERCENTAGE OF CHILDREN WHO INGEST PAINT CHIPS

In a study involving 2,402 children attending the Child Development Center of the University of Virginia, de la Burde and Reames (1973) reported that 9% of mothers of children between eight months and seven years of age responded that their child exhibited pica for paint or plaster. A similar estimate (10%) was reported for 205 children ages 1 to 2 years in the Rochester study (USHUD, 1995a). For this risk analysis, the incidence of paint pica is assumed to be 9% of children living in homes with damaged lead-based paint (defined as greater than 0 ft^2 of interior or exterior deteriorated lead-based paint). Both children with recent paint chip ingestion and those who ingested paint chips at some time are included in the 9%.

Although detailed information on the condition of homes was not available, children in the University of Virginia study were generally from low income families and lived in substandard housing, where flaking paint or falling plaster were likely to be accessible. However, it is not clear whether the homes of all children with pica contained paint chips. Of the children reported to have a history of pica for paint or plaster, 83% lived in urban neighborhoods with old and dilapidated housing and 9% lived in newer urban or suburban homes. The remaining children lived in rural areas, or the type of housing was unknown. It was reported that some children with a history of pica were known to have eaten paint chips or plaster in the home of a relative or babysitter, where they spent a large part of the day. Thus it is possible that children living in homes without damaged lead-based paint may ingest paint chips. It is also possible that children may not be observed eating paint chips, or may ingest paint chips by chewing on intact paint. Because blood-lead concentrations are adjusted only for the incidence of observed pica, only in homes with damaged lead-based paint, the effect of pica on childhood blood-lead levels may be underestimated in the risk analysis. However, it is assumed that the impact is minimal, because estimated blood-lead concentrations are adjusted for pica even in homes with small amounts of damaged lead-based paint.

For HUD National Survey homes where no damaged lead-based paint is present, the IEUBK model and the empirical model (with paint/pica = 0) predicted values are used to estimate

blood-lead concentrations for all children represented by the home. When damaged lead-based paint is present, the same predicted values are used to estimate blood-lead concentrations under each model for 91% of the children, who are assumed not to ingest paint chips. The modeling approaches differ for the remaining 9% of children, who are assumed to ingest paint chips. Because the empirical model incorporates the effect of pica for paint, the model predicted values are used to estimate blood-lead concentrations for children who ingest paint chips. The IEUBK model does not include a direct mechanism for estimating the effect of pica for paint. Thus, adjustments are made to the IEUBK model estimates after the model is applied. The assumptions utilized in this risk analysis, to account for the effect of paint pica under the IEUBK model, are described in the sections that follow.

BLOOD-LEAD CONCENTRATION FOR CHILDREN WITH RECENT PAINT CHIP INGESTION (IEUBK MODEL)

When the IEUBK model is used, the blood-lead concentration is set equal to 63 µg/dL for children who have recently ingested paint chips. The basis underlying this blood-lead concentration and the percentage of children assumed to have recently ingested paint chips are discussed in this section.

The effect of pica for paint will be applied only for HUD National Survey homes where damaged lead-based paint is present. Fifty-five of the 284 homes in the HUD National Survey have damaged lead-based paint. These homes represent 15.2% of U.S. housing, based on 1997-projected weights used in the risk analysis.

Of the 924 children ages 1-2 years in the NHANES III Survey (Brody, et al., 1994), just one child had a blood-lead level greater than 40 µg/dL. The percentage of children ages 1-2 with blood lead greater than 40 µg/dL, adjusted for sampling weights, is 0.03%.

Information on condition of housing was not available for NHANES III participants. It is assumed that blood-lead levels greater than 40 µg/dL are extremely rare in homes with no damaged lead-based paint. Thus the entire 0.03% of children nationwide with blood lead greater than 40 µg/dL are assumed to reside in the 15.2% of homes with damaged lead-based paint. Combining these figures, we estimate that 0.20% of children in homes with damaged lead-based paint have blood-lead levels greater than 40 µg/dL.

A St. Louis study (McElvaine, et al., 1992) found that 13 of 90 (14.4%) children less than age 3 years with blood-lead levels greater than 40 µg/dL, or less than age 7 years with blood lead levels greater than 50 µg/dL, had radiographic evidence of recent paint chip ingestion. This information, combined with the preceding estimate, leads us to conclude that 0.03% of children in homes with damaged lead-based paint have blood lead greater than 40 µg/dL due to recent paint chip ingestion. Table D1-1 shows step by step the methodology for computing the percentage of children living in homes with damaged lead-based paint who have blood-lead levels greater than 40 µg/dL and have recently ingested paint chips. The underlying assumptions of this approach are that 1) blood-lead concentrations are greater than equal to 40 for children who have recently ingested paint chips containing lead and 2) only children who reside in homes with damaged lead

based paint ingest paint chips containing lead. The 13 children in the St. Louis study, who were confirmed to have ingested paint chips, had a mean blood-lead level of 63 µg/dL. The blood lead levels of children with recent pica (0.03% of children in homes with damaged lead-based paint) will be mapped to 63 µg/dL.

Table D1-1. Calculation of Percentage of Children Who Have Recently Ingested Paint Chips.

Variable Name	Variable Definition	Method of Calculation	Value
PC_EAT	Percentage of children with blood lead concentration ≥ 40 µg/dL, living in homes with damaged lead-based paint, who have recently ingested paint chips containing lead.	(PbB ≥ 40 µg/dL| Damaged LBP) * (PC_EAT | PbB ≥ 40 µg/dL)	.197% x .144 = .03%
(PbB ≥ 40 µg/dL| Damaged LBP)	Percentage of children with blood-lead concentration ≥ 40 µg/dL, living in homes with damaged lead based paint.	(PbB ≥ 40 µg/dL) / (Damaged LBP)	0.03% / 0.152 = 0.197%
Damaged LBP	Percent of US housing units with damaged lead based paint.	Percentage of housing units with damaged lead-based paint, estimated in the HUD National Survey.	15.2%
PbB ≥ 40 µg/dL	Percentage of children aged 1-2 with blood-lead concentration ≥ 40 µg/dL.	Taken from NHANES III for children 1-2 years of age.	0.03%
(PC_EAT | PbB ≥ 40 µg/dL)	Percentage of children with blood-lead concentration ≥ 40 µg/dL who have recently ingested paint chips.	Taken from McElvaine's St. Louis study.	13/90 = 14.4%

BLOOD-LEAD CONCENTRATION FOR CHILDREN WHO INGESTED PAINT CHIPS AT SOME TIME (IEUBK)

For HUD National Survey homes with damaged lead-based paint, 9% of the children represented by those homes are assumed to ingest paint chips, with 0.03% of children assumed to have recent paint chip ingestion, as described above. The remaining 8.97% of children are assumed to have ingested paint chips at some time, but not recently. The geometric mean blood-lead concentration for the 8.97% of children in homes with damaged lead-based paint, who have ingested paint chips at some time, is estimated to be 3 µg/dL greater than the IEUBK predicted value for children who do not eat paint chips. The basis for this adjustment is presented in this section.

Although the University of Virginia study was used to estimate the percentage of children who ingest paint chips, children in this study would have been exposed to lead from sources, such as automobile exhaust, no longer present in the environment. Thus their blood-lead levels, if available, would not be comparable to those of present-day children. A current estimate of the effect of pica for paint may be derived from Rochester Lead-in-Dust study (USHUD, 1995a). In that study, 20 of 205 children (10%) were reported to exhibit pica for paint. The geometric mean blood lead for children who were reported to have ingested paint chips was 9.1 µg/dL, while the

geometric mean blood lead for children who were reported to have never ingested paint chips was 6.1 µg/dL. Thus, the geometric mean blood-lead concentration for children who ingested paint chips at some time is assumed to be 3.0 µg/dL greater than the IEUBK model predicted geometric mean for children who do not ingest paint chips.

APPENDIX D2

Results of Three Published Meta-Analyses on the Relationship Between IQ Point Loss and Childhood Blood-Lead Levels

APPENDIX D2

RESULTS OF THREE PUBLISHED META-ANALYSES ON THE RELATIONSHIP BETWEEN IQ POINT LOSS AND CHILDHOOD BLOOD-LEAD LEVELS

INTRODUCTION

The association between blood-lead levels and low IQ scores has been consistently reported in the scientific literature. The estimates of the dose-response relationship published in the literature have been combined via meta-analysis and reported in the three articles listed below. This appendix provides a summary of each article and a discussion of the key results, relative to this risk analysis. The studies cited in these articles are summarized in Tables D2-1 and D2-2 at the end of this appendix.

PRIMARY REFERENCES

Schwartz, J., 1993, Beyond LOEL's, p Values, and Vote Counting: Methods for Looking at the Shapes and Strengths of Associations, Neuro Toxicology 14(2-3):237-246.

Schwartz, J., 1994, Low-Level Lead Exposure and Children's IQ: A Meta-analysis and Search for a Threshold, Environmental Research 65:42-55.

Pocock, S. J., Smith, M., and Baghurst, P., 1994, Environmental Lead and Children's Intelligence: A Systematic Review of the Epidemiological Evidence, BMJ 309:1189-1197.

SUMMARY OF SCHWARTZ, J., 1993

This paper uses examples from the lead literature to illustrate statistical methods for determining the shape of dose-response relationships, including the possible existence of thresholds, and for assessing the strengths of associations within a study and for the literature as a whole. Of interest to this risk analysis is a meta-analysis of the results from 7 studies that estimated a slope for the relationship between children's blood-lead levels and IQ scores. These studies used linear, or log-linear, regression models to fit the relationship between IQ scores and PbB in children. Up to 17 additional covariates were included in the models. The weighted mean regression slope over the 7 studies, weighted by the inverse of the estimated variance, was -0.245 (±0.039). That is, a 1 µg/dL increase in PbB was associated with a 0.245 decrease in IQ score.

SUMMARY OF SCHWARTZ, J., 1994

This article focuses on the relationship between blood lead and IQ scores, while the earlier paper by Schwartz used this relationship to illustrate a statistical method. The 1994 paper presents a meta-analysis of 7 studies, some of which had been cited in the earlier paper, that

estimated a slope for the relationship between children's blood-lead levels and IQ scores. Three longitudinal and four cross-sectional studies were included in the analysis. The studies used linear, or log-linear, regression models to fit the relationship between IQ scores and PbB in children. Additional covariates were included in the models. A random effects model was employed in the meta-analysis, using the method of Dersimonian and Laird (1986). The weighted mean regression slope over the 7 studies, weighted by the inverse of the estimated variance, was -0.257 (±0.041). That is, a 1 µg/dL increase in PbB was associated with a 0.257 decrease in IQ score.

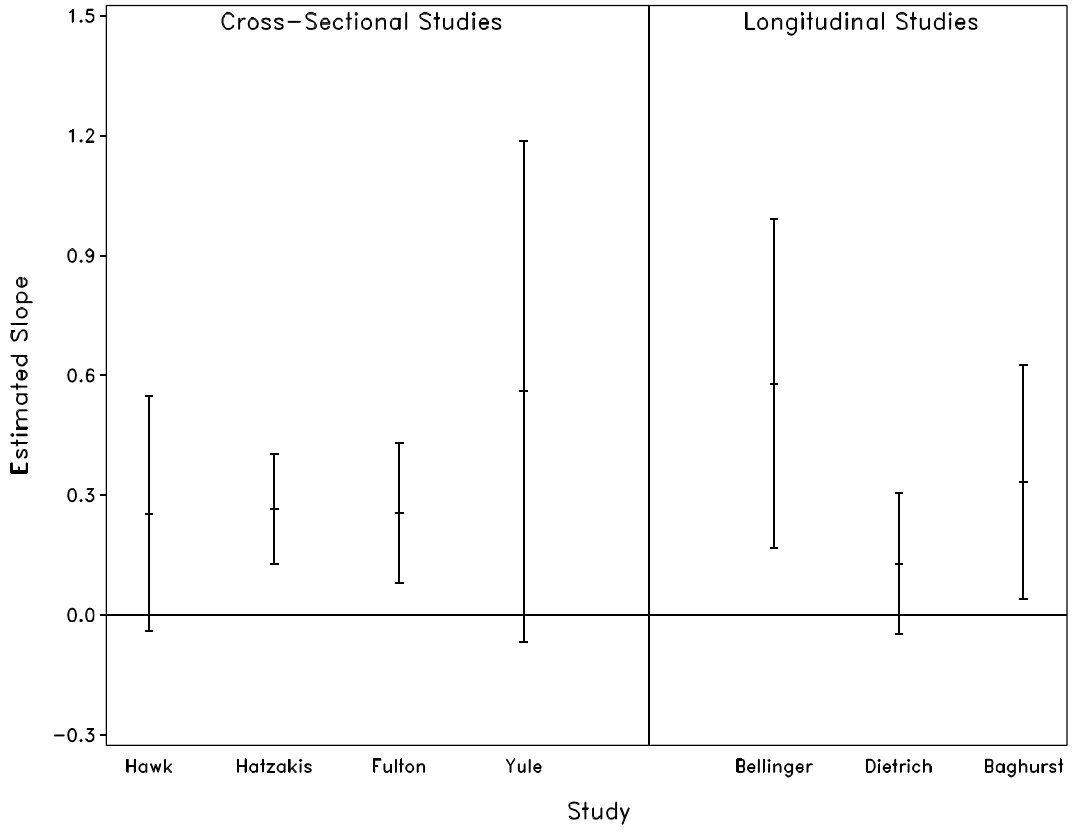


Figure D2-1. Estimated Slopes from the Seven Studies Used in the Schwartz (1994) Meta-analysis, with 95% Confidence Intervals.

SENSITIVITY ANALYSIS

Schwartz conducted a sensitivity analysis to measure the robustness of the meta-analysis and to determine the influence of differences in study design and study populations. The results of the sensitivity analysis are summarized in the following table.

Revised Analysis	Resulting Slope (± 1 standard error of the mean)
Study with Largest Effect Size Removed:	-0.243 (± 0.034)
Study with Most Significant Effect Removed:	-0.252 (± 0.058)
Add 8 Studies with No Effect (each with average weight of the 7 studies):	Association still significant, but slope reduced to about half of original estimate
Longitudinal vs. Cross-sectional:	-0.296 (± 0.125) vs. -0.269 (± 0.051)
Disadvantaged vs. Nondisadvantaged Lifestyle:	-0.185 (± 0.092) vs. -0.289 (± 0.050)
Add 2 Studies that Included Younger Children:	-0.239 (± 0.031)

Three analyses were used to examine the robustness of the meta-analysis. First, the study with the largest effect size (Bellinger et al., 1992) was removed. Next, the study with the most significant effect (Hatzakis et al., 1987) was removed. Based on these results, Schwartz concluded that the meta-analysis was not dominated by any individual study. The third analysis added eight hypothetical studies that reported no association between blood-lead levels and IQ scores. Each study was assigned the average weight of the seven original studies. In this analysis, the association between blood-lead levels and IQ scores was still highly significant ($p<0.01$), but the estimated slope was reduced.

Additional analyses were conducted to determine the effect of differences in study design (longitudinal vs. cross-sectional) and study populations (advantaged vs. disadvantaged, age of child). Schwartz concluded that there was little evidence of a difference in effect size between longitudinal and cross-sectional studies. It did appear that estimates of IQ loss were lower in studies of disadvantaged children. Schwartz suggested that this result may be due to the greater influence of confounding variables in a disadvantaged population. Finally, the addition of two studies that examine younger children did not have a great impact on the estimated slope.

THRESHOLD ANALYSIS

The question of whether a threshold exists in the relationship between IQ scores and PbBs was examined through a meta-analysis that compared studies with different mean blood lead levels. In studies with mean blood lead levels of 15 µg/dL or lower, the estimated slope was -0.323 (±0.126) compared to -0.232 (±0.040) for studies with means above 15 µg/dL. Thus, if anything, a trend toward a higher slope at lower concentrations was observed. This result suggests that the log-linear model may be more appropriate than the linear model, for this relationship.

An alternative approach to the threshold issue examined the data from the Boston study (Bellinger, 1992) more thoroughly. The Boston study was chosen because it had the lowest mean PbB. For this analysis, separate regression models for IQ score and PbB were fit using the same

set of covariates. A nonparametric smoothed curve (LOESS) was fit to the relationship between the two sets of residuals. Based on this analysis, Schwartz concluded that the relationship between blood lead and IQ continues at PbB below 5 µg/dL in this study, i.e., no threshold was evident.

SUMMARY OF POCOCK, S. J., SMITH, M., AND BAGHURST, P., 1994

This paper presents a meta-analysis of 26 epidemiological studies: 5 prospective studies, 14 cross-sectional studies of blood-lead, and 7 cross-sectional studies of tooth-lead. The three types of studies are considered in separate meta-analyses. The results are summarized as follows:

Analysis	Resulting Slope (± 1 standard error of the mean)
Prospective Studies, PbB at Birth:	0.018 (± 0.062)
Prospective Studies, PbB around 2 Years:	-0.185 (± 0.051)
Prospective Studies, Postnatal Mean PbB:	-0.088 (± 0.058)
Cross-Sectional Blood-Lead Studies:	-0.253 (± 0.041)
Cross-Sectional Blood-Lead Studies, Excluding Shanghai:	-0.174 (± 0.043)
Cross-Sectional Tooth-Lead Studies:	-0.095 (± 0.025)

Only the analysis of cross-sectional blood-lead studies had a statistically significant slope.

DISCUSSION

There was considerable overlap in the studies cited by the three meta-analysis papers. Two studies, Fulton et al. (1987) and Yule et al. (1981), were cited in all three papers, while several others were cited in two of the three papers. In addition, some studies cited by Schwartz (1993) or Pocock were used by Schwartz (1994) in the sensitivity analysis.

The three papers are directly comparable in that a common endpoint was used for all meta-analyses. For the meta-analysis endpoint, the regression coefficients and standard errors calculated by the original authors were used to estimate the change in IQ for an increase in blood-lead from 10 to 20 µg/dL. This was necessary, because some of the original authors worked with log-transformed data, while others did not transform the data. In most cases, the regression coefficients were adjusted for other covariates included in the model. The other covariates varied from study to study. For this risk analysis, we have converted the estimated change in IQ back to a slope for untransformed blood-lead concentrations.

The Schwartz (1993) paper focuses on introducing the statistical methods to a non-technical audience. The Schwartz (1994) and Pocock papers focus on the relationship between

IQ and blood-lead levels. The Schwartz (1994) paper includes a sensitivity analysis and search for threshold in the relationship. These topics are not covered in the Schwartz (1993) and Pocock papers. However, in the meta-analysis of prospective studies, the Pocock paper does include separate analyses for blood-lead measures at three ages. Also, one of the studies (Schroeder, 1985) used in the Schwartz (1993) paper included approximately 50 children under 30 months of age. This study and another (Ernhart, 1989) with younger children were included in the sensitivity analysis in Schwartz (1994).

The Pocock paper analyzes longitudinal and cross-sectional studies separately, while the Schwartz papers include both types of studies in the same meta-analysis. The Schwartz (1994) paper considers the study designs separately in the sensitivity analysis. It is important to point out that the measures of blood-lead concentration are different between longitudinal and cross-sectional studies. Cross-sectional studies generally have a single blood lead measurement, taken when the IQ test is administered to school age children. Longitudinal studies generally have several blood-lead measurements available, which may be taken years prior to the IQ testing. In some longitudinal studies (Dietrich et al, 1993; Baghurst et al, 1992), the lifetime average blood-lead concentration is related to IQ. In others (Bellinger et al, 1992; Ernhart et al, 1989), blood-lead concentration at a specified age is related to IQ. The interpretation of the modeled relationships should take into account the differing blood-lead measurements employed. While each author attempts to take this into account, by modeling longitudinal and cross-sectional studies separately, neither distinguishes between the differing measures of blood-lead concentration in longitudinal studies.

In the analysis of prospective studies, Pocock includes an analysis of how PbB at approximately age 2 affects IQ measured at school age. The slope for this analysis (-0.185) is less than the values (approximately -0.25) from Schwartz (1993 and 1994) and the Pocock cross-sectional studies analysis.

Both Schwartz (1994) and Pocock included "full scale IQ score" in school-age children as a selection criteria for studies used in the meta-analysis. Most of the studies cited used the Wechsler Intelligence Scale for Children - Revised (WISC-R) test. The 1993 Schwartz paper includes one study, Schroeder (1985), that uses the Bayley Scales of Infant Development (BSID), for children less than 30 months of age. The BSID score is not directly comparable with the IQ scores, as this test measures developmental endpoints as well as cognitive ability.

Table D2-1. Design Information for Studies that Investigate the Relationship Between Child's IQ and Blood-Lead Level.

Primary References That Cite the Study	Study	Type of Study	Year(s) of Study	Location of Study Participants	Age of Study Participants		IQ Test Instrument	Sample Size	Other Study Information
					Blood Lead Measure	IQ Measure			
Schwartz (1993) Schwartz (1994)	Hatzakis et al. (1987)	Prospective	1985	Lavrion, Greece (a lead smelter city; soil lead levels of 1,300-18,000 ppm)		Primary school age	WISC-R	509	Study participants enrolled in one of four schools in the town in 1984-85.
Pocock	Hatzakis et al. (1989)	Prospective		Lavrion, Greece (a lead smelter city)		6-12 yrs	WISC-R	509	
Schwartz (1993)	Bellinger et al. (1991)	Prospective	Mid- to late-1980s	Boston, MA		Approx. 57 mos	GCI	150	Middle and upper-middle class families, not in inner-city or housing projects. Children born at Brigham and Women's Hospital from 1979-1981
Schwartz (1994) Pocock	Bellinger et al. (1992)	Prospective	1979(Aug.) - 1981(April)	Boston, MA	24 months	School Age	WISC-R	147	Middle class, advantaged
Schwartz (1994) Pocock	Baghurst et al. (1992)	Prospective	1979-1982	Port Pirie, Australia	0 - 3 yrs	7 yrs	WISC-R	494	Smelter town and rural surroundings, middle class families
Pocock	Ernhart et al. (1989)	Prospective		Cleveland, OH	at 2yrs	5 yrs	WPPSI	212	Inner city, disadvantaged, 50% of mothers alcoholic
Pocock	Cooney et al. (1991)	Prospective	1983-1990	Sidney, Australia	1 and 2 yrs	7 yrs	WISC-R	175	Mixed urban
Schwartz (1993)	Schroeder et al. (1985)	Prospective	1977-1978	Wake County, NC		10 mos - 6.5 yrs (half < 30 mos)	BSID (< 30 mos) SBIS (≥ 30 mos)	104	Low income families
Schwartz (1993) Schwartz (1994)	Hawk et al. (1986)	Replication of Schroeder Study		Lenoir & New Hanover counties, NC		3-7 yrs	SBIS	75	Black study participants from low income and SES families, at high risk of exposure to deteriorated LBP
Schwartz (1994) Pocock	Dietrich et al. (1993)	Prospective		Cincinnati, OH	0 - 3 yrs	Approx. 6.5 yrs	WISC-R	231	Inner city, black, disadvantaged

Primary References That Cite the Study	Study	Type of Study	Year(s) of Study	Location of Study Participants	Age of Study Participants		IQ Test Instrument	Sample Size	Other Study Information
					Blood Lead Measure	IQ Measure			
Schwartz (1993) Schwartz (1994) Pocock	Yule et al. (1981)	Pilot Study	Summer 1980 (PbB taken 9-12 months earlier)	Outer London, England		6-12 yrs	WISC-R	166	Results for younger children are reported elsewhere.
Schwartz (1993) Pocock	Lansdown et al. (1986)	Replication of Yule Study		Within 1 km of a factory in London, England		6-12 yrs	WISC-R	166	Mostly middle class families with homes near a main road
Pocock	Winneke et al (1990)	Multi-Center, Cross - Sectional Study		Bucharest		9.2 yrs (mean age)	WISC-Short Form	301	General population
				Budapest		8.5 yrs (mean age)	WISC-Short Form	254	General population
				Moden		7.8 yrs (mean age)	WISC-Short Form	216	Industrial city, lead industry
				Sofia		7.3 yrs (mean age)	WISC-Short Form	142	General population
				Dusseldorf		6.5 yrs (mean age)	WISC-Short Form	109	Industrial city, near smelter
				Dusseldorf		8.3 yrs (mean age)	WISC-Short Form	109	Industrial city, near smelter
Schwartz(1994) Pocock	Silva (1988)	Cross - Sectional	1972-1973	Dunedin, New Zealand		11 yrs (mean age)	WISC-R	579	Mixed urban and rural
Pocock	Harvey et al (1988)	Cross - Sectional	Late 1979-early1981	Birmingham, England		5.5 yrs (mean age)	WPPSI	177	Mixed, inner city
Pocock	Wang et al (1989)	Cross - Sectional		Shanghai, China		6-14 yrs	WISC-R	157	Near battery plant, rural control
Pocock	Winneke et al (1985a)	Cross - Sectional		Nordenham, Germany		7 yrs	WISC-R	122	Smelter town, rural surroundings
Schwartz (1993) Schwartz (1994) Pocock	Fulton et al. (1987)	Cross - Sectional	1983-1985	Edinburgh, Scotland		6-9 yrs	BAS	501	Study participants enrolled in one of 18 primary schools Mixed urban - previously hish water lead

Table D2.2. Summary of Results from Studies that Investigate the Relationship Between Child's IQ and Blood Lead Level

Table D2-2. Summary of Results from Studies that Investigate the Relationship Between Child's IQ and Blood-Lead Level. (Continued)

Primary References That Cite the Study	Study	PbB of Study Participants[1] (µg/dL)		IQ of Study Participants[2]		Measure of Association Between IQ and Blood-Lead Levels[3]			
		Range	Summary Statistics	Endpoint Type	Range/Summary Statistics	Measure	P-Value	Covariates	Other Information
Schwartz (1993) Schwartz (1994)	Hatzakis et al. (1987)	7.4 - 63.9	AM = 23.7 STD = 9.2 10%ile = 13.9 50%ile = 21.5 90%ile = 36.0	WISC-R		-0.270 change in IQ per unit increase in PbB (-0.403, -0.137)	< 0.001	17 potential confounders or IQ correlates[4] (called the "optimal" model)	Dose-response investigation showed no PbB effect on IQ when PbB < 25 µg/dL.
Pocock	Hatzakis et al. (1989)	7.4-63.9	AM = 23.7 STD = 9.2		AM = 87.7 STD = 14.8	-2.7 change in IQ for increase from 10-20 µg/dL in PbB	< 0.001	Up to 24, including mother's IQ	Dose-reponse curve showed evidence of a threshold at the level of about 25 µg/dL PbB
Schwartz (1993)	Bellinger et al. (1991)	0.0 - 23.3	AM = 6.4 STD = 4.1 19% were > 10µg/dL 4% were > 15µg/dL	GCI	80-150 AM = 115.5 STD = 14.5	-2.28 change in IQ per unit increase in Log(PbB) (-6.0, 1.4) --- -0.250 change in IQ per unit increase in PbB from 5-15 µg/dL PbB	0.23	13 covariates[5]	Regression diagnostics were used to check the robustness of estimates. These results reflect only PbB data at age 57 months.
Schwartz (1994) Pocock	Bellinger et al (1992)		AM = 6.5 STD = 4.9	WISC-R	71-147 AM = 119.1 STD = 14.8	-5.8 change in IQ for increase from 10 to 20 µg/dL in PbB	0.007	HOME mother's IQ, 8 other covariates[8]	Slightly elevated blood lead levels around the age of 24 months are associated with intellectual and academic performance deficits at age 10 years.
Schwartz (1994) Pocock	Baghurst et al (1992)		AM = 20	WISC-R	AM = 104.7	-3.3 change in IQ for an increase from 10-20 µg/dL in PbB	0.04	HOME, mother's IQ, 11 others [9]	Found low-level exposure to lead during early childhood is inversely associated with neuropsychological development through first seven years of life.
Pocock	Ernhart et al (1989)		AM = 16.7 STD = 6.45	WPPSI	AM = 87.5 STD = 16.6	-1.1 change in IQ for an increase from 10-20 µg/dL in PbB	< 0.01	HOME , mothers IQ, and 11 others[11]	
Pocock	Cooney et al (1991)		AM = 14.2	WISC-R		0.39 change in IQ for an increase from 10-20 µg/dL in PbB		HOME ,mothers IQ, and 4 others [12]	

Table D2-2. Summary of Results from Studies that Investigate the Relationship Between Child's IQ and Blood-Lead Level. (Continued)

Primary References That Cite the Study	Study	PbB of Study Participants[1] (µg/dL)		IQ of Study Participants[2]		Measure of Association Between IQ and Blood-Lead Levels[3]			
		Range	Summary Statistics	Endpoint Type	Range/Summary Statistics	Measure	P-Value	Covariates	Other Information
Schwartz (1993)	Schroeder et al. (1985)	6 - 58		BSID (< 30 mo.) SBIS (≥30 mo.)	45-140	-0.199 change in IQ per unit increase in PbB	< 0.01	7 covariates[6] plus interaction with PbB. Quadratic and cubic components of PbB also considered.	Unforced stepwise regression. SES was only other significant covariate.
Schwartz (1993) Schwartz (1994)	Hawk et al. (1986)	6.2 - 47.4	AM = 20.9 STD = 9.7	SBIS	59-118	-0.255 change in IQ per unit increase in PbB (-0.554, 0.043)	< 0.05	Gender, HOME score, maternal IQ	
Schwartz (1994) Pocock	Dietrich et al (1993)		AM = 15.2 STD = 11.3	WISC-R	AM = 86.9 STD = 11.3	1.3 esimated loss in IQ for an increase from 10 to 20 µg/dL in PbB	< 0.10	HOME score, maternal IQ, birth weight, birth length, child sex, cigarette consumption during pregnancy	Postnatal PbB concentrations were inversely associated with Full Scale IQ.
Schwartz (1993) Schwartz (1994) Pocock	Yule et al. (1981)	7 - 33	AM = 13.52 STD = 4.13 80% were > 10µg/dL 4.8% were > 20µg/dL	WISC-R	AM = 98.21 STD = 13.44	-8.08 change in IQ per unit increase in Log(PbB) (4.63) -0.560 change in IQ per unit increase in PbB from 10-20 µg/dL	0.084	Age, social class	Social class was considered a crude measure.
Schwartz (1993) Pocock	Lansdown et al. (1986)	7 - 24	AM = 12.75 STD = 3.07 77% were > 10µg/dL 1.5% were > 20µg/dL	WISC-R WISC-R	AM = 105.24 STD = 14.20	2.15 change in IQ per unit increase in Log(PbB) 0.149 change in IQ per unit increase in PbB from 10-20 µg/dL	0.63	Age, social class	N = 86 for regression analysis. Social class was also a significant factor.

Table D2-2. Summary of Results from Studies that Investigate the Relationship Between Child's IQ and Blood-Lead Level. (Continued)

Primary References That Cite the Study	Study	PbB of Study Participants[1] (µg/dL)		IQ of Study Participants[2]		Measure of Association Between IQ and Blood-Lead Levels[3]			
		Range	Summary Statistics	Endpoint Type	Range/Summary Statistics	Measure	P-Value	Covariates	Other Information
Pocock	Winneke et al (1990) Bucharest		GM = 18.9 STD = 1.3	WISC-Short Form			<0.1	Gender, age, social class, mother's education	
	Winneke et al (1990) Budapest		GM = 18.2 STD = 1.7	WISC-Short Form			< 0.1	Gender, age, social class	
	Winneke et al (1990) Moden		GM = 11.0 STD = 1.3	WISC-Short Form			< 0.1	Gender, age, social class, mother's education	
	Winneke et al (1990) Sofia		GM = 18.2 STD = 1.6	WISC-Short Form			< 0.1	Gender, age, social class, mother's education	
	Winneke et al (1990) Dusseldorf		GM = 8.3 STD = 1.4	WISC-Short Form	AM = 116		< 0.1	Gender, age, social class, mother's education	
	Winneke et al (1990) Dusseldorf		AM = 7.4 STD = 1.3	WISC-Short Form			< 0.1	Gender, age, social class, mother's education	
Schwartz (1994) Pocock	Silva (1988)	4 - 50 µg/dL	AM = 11.1 STD = 4.91	WISC-R	AM = 108.9 STD = 15.12	Loss of 1.51 in IQ for an increase in PbB of 10-20µg/dL		None	
Pocock	Harvey et al (1988)	0.2-1.4 mol/L	AM = 12.3 STD = 0.2	WPPSI	AM = 105.9 STD = 10.6			None	No significant relationship was found between overall IQ and PbB

Table D2-2. Summary of Results from Studies that Investigate the Relationship Between Child's IQ and Blood-Lead Level. (Continued)

Primary References That Cite the Study	Study	PbB of Study Participants[1] (µg/dL)		IQ of Study Participants[2]		Measure of Association Between IQ and Blood-Lead Levels[3]			
		Range	Summary Statistics	Endpoint Type	Range/Summary Statistics	Measure	P-Value	Covariates	Other Information
Pocock	Wang et al (1989)	4.5 - 52.8 µg/dL	AM = 21.1 STD = 10.11	WISC	AM = 89	A decrease of IQ of 9 per 10µg/dL increase in PbB		Mother's education and 4 others [10]	Found a dose - effect relation between PbB and IQ even after confounding variables were controlled for by stepwise regression analysis
Pocock	Winneke et al (1985a)	4.4 - 23.8 µg/dL	AM = 8.2 STD = 1.4	WISC-R	AM = 120.2 STD = 10.3		< 0.1	Age, sex and hereditary background	
Schwartz (1993) Schwartz (1994) Pocock	Fulton et al. (1987)	3.3 - 34	GM = 11.5 1.2% were > 25µg/dL	BASC	AM = 112 STD = 13.4	-3.70 change in IQ per unit increase in Log(PbB) (1.31) --- -0.256 change in IQ per unit increase in PbB from 10-20 µg/dL	0.003	13 covariates[7] + school attended ("optimal" regression model)	Adjusted R^2 = 45.5%

Notes for Table D2-2:

(1) "Range" indicates the observed range of PbB levels among the study participants. Among the summary statistics, AM = arithmetic mean; GM = geometric mean; STD = standard deviation; x%ile = x percentile of observed distribution.

(2) "Type" indicates the type of IQ endpoint measured in the study. WISC-R = Wechsler Intelligence Scale for Children - Revised (full-scale IQ measurement); GCI = McCarthy Scales of Children's Abilities: General Cognitive Index; BSID = Bayley Scales of Infant Development; SBIS = Stanford-Binet Intelligence Scale; BASC = British Ability Scales: Combined Score. Among the summary statistics, AM = arithmetic mean; STD = standard deviation.

(3) Results are the outcome of a regression analysis to predict IQ endpoint based on PbB level and other covariates. "Measure" is the estimated slope parameter indicating the change in IQ measurement associated with a unit change in the (possibly transformed) PbB level. If the PbB level is transformed, the change in IQ measurement over a given range of the untransformed PbB level is also given. When available, a 95% confidence interval associated with the slope estimate is given, or a standard error associated with the estimate. "P-value" is for the test that the slope parameter is equal to zero versus an alternative that it is not zero. "Adjusted covariates" indicates the number of covariates included in the regression model; these covariates are named if the number is small. "Other information" indicates specifics associated with the regression fit (e.g., method used, whether a log-transformation was taken on the PbB level prior to analysis, information on the covariates).

(4) Covariates include parental IQ, birth order, family size, father's age, parental education, alcoholic mother, age, bilingualism, birth weight, length of child's hospital stay after birth, walking age, history of CNS disease, history of head trauma, illness affecting sensory function, parent's divorce.

(5) Covariates include family social class, material IQ, preschool attendance, HOME total score, # hours per week of "out-of-home" care, # changes in family residence since birth, medication use in preceding month, # adults in household, gender, race, birth weight, material marital status, birth order.

(6) HOME score, maternal IQ, child's age, child's sex, SES of parents, type of IQ test, presence of father in home, number of siblings.

(7) Parent's vocabulary and matrices tests, child's interest score, age, father's qualifications, length of gestation, parental involvement with school score, class year, # days absent from school, sex, standardized height, car/telephone ownership, employment status of father.

(8) Child stress, maternal age, race, SES, sex, birth order, martial status, number of residence changes prior to age 57 months

(9) Sex, parents' level of education, maternal age at delivery, parents' smoking status, socio-economic status, quality of the home environment, birth weight, birth order, feeding method (breast feeding, bottle, or both), duration of breast - feeding, and whether the child's natural parents were living together

(10) Age, sex, father's education, father's occupation, father's daily smoking quantity

(11) Sex, race, birth weight, birth order, gestational age at birth, parental education, maternal variables like PPVT-R, AFI, MAST SCORE, AA/day in pregnancy, cigarettes per day, and use of marijuana and other drugs in pregnancy, medical problems and psychosocial problems.

(12) Gestational age, education of the mother, education and occupational status of the father.

APPENDIX E1

Methodology for Estimating Health Effects From Blood-Lead Distribution

APPENDIX E1

METHODOLOGY FOR ESTIMATING HEALTH EFFECTS FROM BLOOD-LEAD DISTRIBUTION

This appendix describes the procedure used in this report for calculating health and blood-lead endpoints for the nation's children aged 1-2 years, based on a distribution of blood-lead concentrations assumed to be lognormal. In this section, GM represents the geometric mean and GSD represents the geometric standard deviation of the blood-lead concentrations.

a. P[PbB > X], where X=10 µg/dL or 20 µg/dL

Because it is assumed that the blood-lead concentration distribution is lognormally distributed, the probability of observing a blood-lead concentration greater than X is expressed as

$$P[PbB > X] = 1 - \Phi\left(\frac{\ln(X) - \ln(GM)}{\ln(GSD)}\right) \quad (1)$$

where $\Phi(z)$ is the probability of observing a value less than z under the standard normal distribution. Therefore, setting X=10 and X=20 in equation (1) will provide estimates of the probability of observing a blood-lead level exceeding 10 µg/dL and 20 µg/dL, respectively.

b. P[IQ < 70]

As indicated in Table E1-1, the estimated probability that a child will have an IQ score less than 70 given the child's blood-lead concentration (PbB) is expressed as a piecewise linear function of PbB. To estimate the probability that a child in the national population has an IQ score less than 70, the blood-lead distribution is used with the information in Table E1-1. Using the notation x_i, α_i, and β_i (i=1,...,10) introduced in the column headings in Table E1-1, and letting LGM = ln(GM) and LGSD = ln(GSD), the expected value of the probability of observing an IQ score less than 70 is

$$P[IQ < 70] = \sum_{i=1}^{10} \alpha_i \cdot \left[\Phi\left(\frac{\ln(x_i) - LGM}{LGSD}\right) - \Phi\left(\frac{\ln(x_{i-1}) - LGM}{LGSD}\right)\right] + K \cdot \sum_{i=1}^{10} \beta_i \cdot \left[\Phi\left(\frac{\ln(x_i) - LGM - (LGSD^2)}{LGSD}\right) - \Phi\left(\frac{\ln(x_{i-1}) - LGM - (LGSD^2)}{LGSD}\right)\right] \quad (2)$$

Table E1-1. Formulas for Estimating the Probability of Observing IQ Score Less Than 70, Given a Child's Blood-Lead Concentration (PbB).

Interval # (i)	Range of PbB (µg/dL) (x_{i-1} < PbB x_i)	Function for Estimating Increased Percentage of Children Having IQ Scores less than 70 (IQ < 70 = α_i + β_i * PbB)
1	0 < PbB 5	IQ < 70 = 0.080 + 0.0036 * PbB
2	5 < PbB 7.5	IQ < 70 = 0.022 + 0.0152 * PbB
3	7.5 < PbB 10	IQ < 70 = -0.152 + 0.0384 * PbB
4	10 < PbB 12.5	IQ < 70 = -0.084 + 0.0316 * PbB
5	12.5 < PbB 15	IQ < 70 = 0.016 + 0.0236 * PbB
6	15 < PbB 17.5	IQ < 70 = -0.260 + 0.0420 * PbB
7	17.5 < PbB 20	IQ < 70 = -0.281 + 0.0432 * PbB
9	20 < PbB 22.5	IQ < 70 = -0.145 + 0.0364 * PbB
9	22.5 < PbB 25	IQ < 70 = -0.532 + 0.0536 * PbB
10	25 < PbB	IQ < 70 = -0.162 + 0.0388 * PbB

Derived From: Wallsten, T.S., and Whitfield, R.G. "Assessing the Risks to Young Children of Three Effects Associated with Elevated Blood-lead Levels." *Report by Argonne National Laboratory*. Report No. ANL/AA-32. Sponsored by the U.S. EPA Office of Air Quality Planning and Standards. 1986.

where K = exp(LGM + (LGSD)^2 /2) and Φ(z) is the probability of observing a value less than z under the standard normal distribution. In calculating (2) use the following [1] conventions: ln (0)=- , ln( )= , Φ(-)=0, and Φ( )=1.

c. P[IQ decrement > x] for x=1, 2, 3

It is assumed that each µg of lead per dL of blood corresponds to a 0.257 decline in IQ score (see Section 4.4 of the §403 Risk Assessment report). Therefore, an IQ decrement exceeding 1 is associated with blood-lead concentrations exceeding 1/0.257 = 3.9 µg/dL. Similarly, blood-lead concentrations exceeding 2/0.257 = 7.8 µg/dL are associated with an IQ decrement exceeding 2, and concentrations exceeding 3/0.257 = 11.7 µg/dL are associated with an IQ decrement exceeding 3. Therefore,

P[IQ decrement > 1] = P[PbB > 3.9 µg/dL]
P[IQ decrement > 2] = P[PbB > 7.8 µg/dL]
P[IQ decrement > 3] = P[PbB > 11.7 µg/dL]

[1] Equation (2) is equivalent to $\sum_{i=1}^{10} \int_{x_{i-1}}^{x_i} (\alpha_i + \beta_i x)\, \phi(x)\, dx$ where $\phi(x)$ is the probability density function of the lognormal distribution with parameters LGM and LGSD.

where the right-hand side of each of these equations is calculated using equation (1) with X=3.9, 7.8, or 11.7.

d. Average IQ points lost (and associated standard deviation)

The (arithmetic) average IQ points lost in the population of children aged 1-2 years is calculated using the properties of the lognormal distribution. If X corresponds to a child's blood-lead concentration and Y is the associated decline in IQ for the child due to the presence of the blood-lead, then it is assumed in this risk assessment that Y = 0.257*X. As X is assumed to be lognormally distributed, it can be shown that Y is also lognormally distributed. Furthermore, an estimate of the expected value of Y (average # IQ points lost) is as follows:

$$\text{Avg. \# IQ points lost} = 0.257 * GM * \exp(\ln(GSD)^2/2) \quad (4)$$

Note that if 0.257 is excluded from the formula in equation (4), the result would be the arithmetic average associated with the distribution of blood-lead concentrations.

APPENDIX E2

Generating Distribution of Blood-Lead Concentrations Based on Model-Predicted Geometric Mean and Geometric Standard Deviation

APPENDIX E2

GENERATING DISTRIBUTION OF BLOOD-LEAD CONCENTRATIONS BASED ON MODEL-PREDICTED GEOMETRIC MEAN AND GEOMETRIC STANDARD DEVIATION

This section discusses how the geometric mean blood-lead concentrations predicted by either model at each housing condition were combined to characterize the national distribution of children's blood-lead concentrations for children aged 1-2. This approach was used for characterizing both pre- and post-intervention distributions predicted with the models.

Historical data suggest that blood-lead concentrations usually follow a lognormal distribution. A lognormal distribution can be characterized using two parameters, the geometric mean which is a measure of the "center" of the distribution, and the geometric standard deviation (GSD) which is a measure of the spread of the distribution. The empirical and IEUBK models both predict a geometric mean blood-lead concentration for a population of children exposed to specific levels of environmental lead. However, a population of children exposed to the same levels of environmental lead would not all have the same blood-lead concentration represented by the predicted geometric mean. Their blood-lead concentrations will vary about the predicted geometric mean because of the many other factors that contribute to children's blood-lead concentrations. These factors include differences in children's activity patterns, tendency to ingest dust or soil, parental supervision, dietary lead, other lead exposures, and amount of lead absorbed due to various biological factors.

Extant data from various studies indicate that the inherent variability in blood-lead concentration among children exposed to similar environmental-lead levels corresponds to a GSD of 1.6, the default GSD recommended in the IEUBK guidance manual (USEPA, 1994a). Under the assumption that blood-lead concentrations have a lognormal distribution with a geometric mean, GM, and GSD of 1.6, the logarithms of the blood-lead concentrations have a normal distribution with mean $\mu = \ln(GM)$ and standard deviation $s = \ln(1.6)=0.47$.

The predicted national distribution of children's blood-lead concentrations was also assumed to follow a lognormal distribution. The predicted geometric mean of the national distribution of children's blood-lead concentrations is calculated by taking a weighted geometric mean of the empirical or IEUBK model-predicted blood-lead concentrations associated with each home in the HUD National Survey, using the HUD National Survey weights adjusted for 1997 population totals.

The predicted national GSD is calculated by taking the square root of the sum of the predicted between-house variability and the assumed within-house variability. Between-house variability represents the variability among the predicted blood-lead concentrations for homes in the HUD National Survey and is computed as the weighted geometric variance of the model predicted blood-lead concentrations for each home in the HUD National Survey, using the adjusted weights for 1997. Within-house variability, variability in blood-lead concentrations of

children exposed to the same levels of environmental lead, is calculated as the weighted mean of the log variances assigned to each HUD National Survey home. This variability is assumed to be characterized by a GSD of 1.6 (log variance = $\ln (1.6)^2$) for all HUD National Survey homes. There is one exception where the child is assumed to have a blood-lead level of 63 µg/dL in which case the GSD is assumed to be 1.

The methodology for characterizing the national blood-lead distribution is slightly different for the IEUBK and empirical models because of the different ways the two models incorporate paint pica. For the empirical model, the national distribution of blood-lead concentrations is characterized as follows. For each house in the HUD National Survey, let N_i be the number of children aged 1-2 years associated with the housing unit, $GM1_i$ denote the model-predicted geometric mean blood-lead concentration for children without pica tendencies, and $GM2_i$ denote the model-predicted geometric mean blood-lead concentration for children with pica tendencies. Recall that $GM2_i$ is calculated only for units containing deteriorated or damaged lead-based paint.

The distribution of children's blood-lead concentrations in homes with no deteriorated lead-based paint was assumed to have a lognormal distribution with geometric mean $GM1_i$ and a GSD of 1.6. Children in housing units with damaged or deteriorated lead-based paint in either the interior or exterior, were partitioned into two groups:

Group #1: Assumed to contain 91% of the children, representing children who show no tendency toward paint pica. The blood-lead concentration distribution of this group is assumed to be lognormal with geometric mean $GM1_i$ and a GSD of 1.6.

Group #2: Assumed to contain 9% of the children representing children who have exhibited some tendency towards paint pica. The distribution of blood-lead concentrations for this group is assumed to be lognormal with geometric mean $GM2_i$ and a GSD of 1.6.

Let N be the sum of N_i across all homes represented in the HUD National survey (i.e., the total number of children aged 1-2 years in the 1997 housing stock). Furthermore, let A denote all housing units in the section containing no deteriorated lead-based paint, and let B denote the housing units that have some deteriorated lead-based paint. Then the aggregated log-transformed geometric mean blood-lead concentration, denoted by µ, is calculated as:

$$\mu = \frac{\left(\sum_{i\ A} N_i * \ln(GM1_i)\right) + \left(\sum_{i\ B} N_i * (0.91 * \ln(GM1_i) + 0.09 * \ln(GM2_i))\right)}{N}$$

The aggregated log-transformed GSD, denoted by s, is calculated as:

$$s = \sqrt{\frac{\left(\sum_{i\ A} K1_i\right) + \left(\sum_{i\ B} (0.91 * K1_i + 0.09 * K2_i)\right)}{N-1} + (\ln(1.6))^2}$$

where $K1_i = N_i*(\mu - \ln(GM1_i))^2$ and $K2_i = N_i*(\mu - \ln(GM2_i))^2$. The resulting national distribution of blood-lead concentrations is assumed to be lognormally distributed with geometric mean equal to e^{μ} and GSD equal to e^{s} .

For the IEUBK model, let GM_i be the model-predicted geometric mean blood-lead concentration for the i^{th} housing unit. For units without any damaged or deteriorated lead-based paint then the distribution of blood-lead concentrations is assumed to be lognormal with geometric mean GM_i and a GSD of 1.6. Children in housing units with damaged or deteriorated lead-based paint in either the interior or exterior, were partitioned into three groups:

Group #1: Assumed to contain 91% of the children, representing those children who show no tendency toward paint pica. The blood-lead concentration distribution of this group is assumed to be lognormal with geometric mean GM_i and a GSD of 1.6.

Group #2: Assumed to contain 8.97% of the children, representing those children who exhibit paint pica, but have not recently ingested Lead-based paint. The distribution of blood-lead concentrations for this group is assumed to be lognormal with geometric mean $GM_i + 3$ and a GSD of 1.6.

Group #3: Assumed to contain 0.03% of the children, representing those children who have recently ingested lead-based paint. The distribution of blood-lead concentrations for this group is assumed to be lognormal with geometric mean 63 µg/dL and a GSD of 1.

The national log-transformed geometric mean blood-lead concentration is:

$$\mu = \frac{\left(\sum_{i\ A} N_i * \ln(GM_i)\right)}{N} + \frac{\left(\sum_{i\ B} N_i * (0.91 * \ln(GM_i) + 0.0897 * \ln(GM_i + 3) + 0.0003 * \ln(63))\right)}{N}$$

The national log-transformed GSD is:

$$s = \sqrt{\frac{\left(\sum_{i\ A} K1_i\right) + \left(\sum_{i\ B} (0.91 * K1_i + 0.0897 * K2_i + 0.0003 * K3_i)\right)}{N - 1}} + V$$

where

$$V = \frac{\left(\sum_{i\ A} N_i * \ln(1.6)^2\right) + \left(\sum_{i\ B} N_i * 0.9997 * \ln(1.6)^2\right)}{N},$$

where $K1_i = N_i*(\mu - \ln(GM_i))^2$, $K2_i = N_i*(\mu - \ln(GM_i + 3))^2$ and $K3_i = N_i*(\mu - \ln(63))^2$. The resulting national distribution of blood-lead concentrations predicted by the IEUBK model is assumed to be lognormally distributed with geometric mean equal to e^{μ} and GSD equal to e^{s}.

APPENDIX F1

Methodology for Estimating Post-Intervention Distribution of Children's Blood-Lead Concentrations Resulting from Proposed §403 Rules

APPENDIX F1

METHODOLOGY FOR ESTIMATING POST-INTERVENTION DISTRIBUTION OF CHILDREN'S BLOOD-LEAD CONCENTRATIONS RESULTING FROM PROPOSED §403 RULES

This appendix details the procedures used to estimate the national distribution of blood-lead (PbB) concentrations in children aged 1-2 years in 1997 immediately after performing the relevant intervention strategies on the nation's housing stock under the proposed §403 rules.

Outline of the Methodology

This methodology characterizes the pre-§403 blood-lead distribution for children aged 1-2 years using reported information from NHANES III. A model-based procedure (either the empirical or IEUBK model) is used to characterize the distribution of blood-lead concentrations at both pre-§403 and post-§403, and the observed differences between the two distributions are identified. Then, a post-§403 distribution that is comparable to the pre-§403 NHANES III distribution is derived based on the differences between the two model-based estimates and the pre-§403 NHANES III distribution.

The methodology consists of the following four steps:

#1. Use blood-lead concentration data reported in the NHANES III to estimate the geometric mean (GM) and the geometric standard deviation (GSD) associated with the baseline (i.e., pre-§403) distribution of blood-lead concentration for children aged 1-2 years.

#2. Use the environmental-lead levels for HUD National Survey units as input to either the IEUBK or Empirical model to obtain a model-based estimate of the geometric mean and the geometric standard deviation (GSD) associated with the baseline distribution of blood-lead concentration for children aged 1-2 years.

#3. Use adjusted (post-§403) environmental-lead levels for HUD National Survey units as input to the model used in Step #2 to estimate the geometric mean and the geometric standard deviation (GSD) associated with the post-§403 distribution of blood-lead concentration for children aged 1-2 years.

#4. Combine the parameters of the three distributions described in #1, 2, and 3 to estimate the geometric mean and GSD of a post-§403 blood-lead distribution that is consistent with the pre-§403 NHANES III distribution determined in Step #1 and the changes in the blood-lead distributions estimated in Steps #2 and #3.

Details of the Methodology

A key assumption in this methodology is that blood-lead concentrations are assumed to be lognormally distributed, regardless of whether they represent pre- or post-§403 concentrations or whether the distribution is based on NHANES III data or is model-based. With this assumption and by estimating the geometric mean and GSD of the distribution, the entire distribution is characterized.

All four steps of the methodology are now discussed in detail.

#1. Use NHANES III to characterize the pre-§403 distribution.

A weighted geometric mean and weighted geometric standard deviation of the blood-lead concentrations are calculated for 1-2 year old children based on NHANES III. The weights are those discussed in Section 3.4.1. Call these variables GM_1 and GSD_1, respectively. These values were calculated as geometric mean (GM_1) = 3.14 µg/dL and geometric standard deviation (GSD_1) = 2.09.

#2. Derive a model-based characterization of the pre-§403 distribution.

Because interventions under §403 have not yet occurred, precluding post-§403 blood-lead concentrations from being directly measured, the blood-lead distribution resulting from the proposed §403 rules must be estimated. For this reason, this methodology characterizes pre- and post-§403 blood-lead distributions that are model-based (i.e., predicted blood-lead concentrations as a function of environmental-lead levels are obtained using either the IEUBK or empirical model).

Environmental-lead levels in the HUD National Survey database are used as input to the model to characterize the pre-§403 distribution of blood-lead in children aged 1-2 years. The model-based pre-§403 blood-lead distribution is assumed to be lognormally distributed. A weighted geometric mean and weighted geometric standard deviation of these concentrations are calculated, where the weights correspond to the number of children associated with each concentration. Call these variables GM_2 and GSD_2, respectively.

#3. Derive a model-based characterization of the post-§403 distribution.

The same method used in Step #2 is used to characterize a model-based post-§403 distribution (Step #3). Step #3 differs from Step #2 in that the environmental-lead levels from the HUD National Survey are adjusted to reflect the effects of intervention. This adjustment is documented in Table 6-2 of Volume I. Let GM_3 and GSD_3 be the weighted geometric mean and geometric standard deviation, respectively, of the predicted post-§403 blood-lead concentrations. Thus, the model-based post-§403 blood-lead distribution is characterized as lognormally distributed with geometric mean GM_3 and geometric standard deviation GSD_3.

#4. Derive a post-§403 distribution from NHANES III and Steps #2 and #3.

The three distributions calculated in Steps #1 through #3 are used to characterize a post-§403 blood-lead distribution that is directly comparable with the pre-§403 distribution determined in Step #1. This distribution is assumed to be lognormal with geometric mean GM_4 and geometric standard deviation GSD_4 calculated by the following formulas:

$$GM_4 = GM_1 * (GM_3 / GM_2) \quad (1)$$

$$GSD_4 = GSD_1 * (GSD_3 / GSD_2) \quad (2)$$

APPENDIX F2

Estimation of Primary Prevention Efficacy Using Model of Bone-Lead Mobilization

APPENDIX F2

ESTIMATION OF PRIMARY PREVENTION EFFICACY USING MODEL OF BONE-LEAD MOBILIZATION

Though the scientific literature documents the effectiveness of a range of behavioral and environmental intervention strategies on their ability to reduce childhood lead exposure, efficacy is measured only among already exposed children (USEPA, 1995b). Specifically, declines in children's blood-lead concentration on the order of 25% as measured 6 to 12 months following a variety of intervention strategies were reported (Copley, 1983; Charney, et al., 1983; Amitai, et al., 1991; Weitzman, et al., 1993; Staes, et al., 1994; Kimbrough, et al., 1994). This secondary prevention intervention effectiveness is likely not representative of the effectiveness being sought from the promulgation of §403. The §403 standards for lead in dust, soil, and paint are mostly intended to prevent childhood lead exposure before it occurs and, therefore, their effectiveness will be assessed by measures of primary prevention efficacy.

Secondary prevention efficacy results are not necessarily representative of those expected from primary prevention because lead present in blood is a combination of current environmental exposure and internal sources of lead. A significant internal source of lead is bone tissue. After prolonged exposure to lead, bone tissue retains much more lead than the other body tissues (Schroeder and Tipton, 1968; Barry and Mossman, 1970; Barry, 1975; Barry, 1981; Leggett, et al., 1982). Nordberg, et al. suggest that bone can become an internal source of lead during periods of reduced external exposure to lead; see also (Rabinowitz, et al., 1976; Barry, 1981; Hyrhorczuk, et al., 1985; Rabinowitz, 1991). The reported declines in blood-lead concentration, therefore, may underestimate the primary prevention effectiveness of the associated intervention strategy.

Unfortunately, there is limited empirical evidence regarding the extent to which bone-lead stores are able to keep blood-lead levels elevated following an intervention, especially concerning children. One study (Markowitz, et al., 1993) measured bone-lead levels in children before and after an intervention, but found no significant decline in the levels over a period of six weeks. Despite the lack of studies concerning children, Nordberg, et al. claim that "skeletal turnover is highest among children under 10 years of age." Several studies have been conducted to study bone-lead mobilization in adults (Rabinowitz, et al., 1976; Hyrhorczuk, et al., 1985; Wrenn, et al., 1972; Cohen, et al., 1973; Rabinowitz, et al., 1973; Batschelet, et al., 1979; Heard, et al., 1984; Marcus, 1985; Christofferson, et al., 1986; Cristy, et al., 1986; Schutz, et al., 1987; Bert, et al., 1989; Nilsson, et al. 1991; Gulson, et al., 1995). For example, Gulson, et al. show that 45% to 70% of lead in the blood of adult women comes from long-term tissue stores, primarily the bone tissue. A similar result was observed in another study on five adult subjects undergoing knee and hip replacement (Smith, et al., 1996).

If the contribution of mobilized bone-lead stores can be characterized, however, it would be possible to translate the documented secondary prevention results into estimated primary prevention results. An approach is presented here for estimating the efficacy of a primary prevention intervention given an observed effectiveness for a secondary prevention intervention.

The approach is based on a bone-lead mobilization model developed to estimate the degree to which bone-lead stores could mask the full effectiveness of an intervention by mobilizing into the child's blood. This model is extensively discussed and its basis documented elsewhere (Rust, et al., 1996), though a summary is provided below.

A Model for Bone-Lead Mobilization

To evaluate the potential for continuing elevated blood-lead levels due to bone-lead mobilization, a two-compartment model (see Figure F2-1) was adopted for the transfer of lead between the blood and bone tissues within the body and elimination of lead from the body.

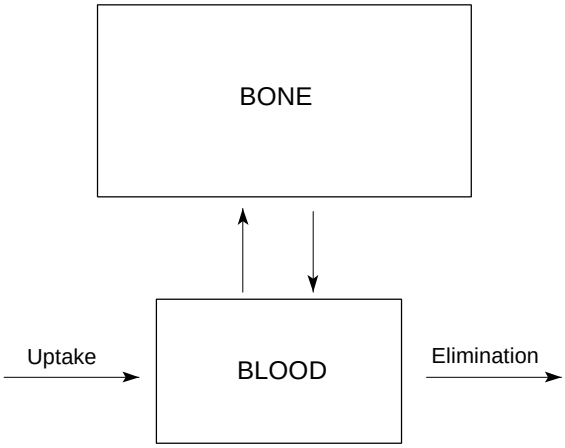


Figure F2-1. Two Compartment Model of Bone-Lead Mobilization.

In this model, lead is taken into the body (from the gastrointestinal tract and lungs) via the blood, transfers between the blood and bone tissue, and is eliminated from the body via the blood. It is assumed that the transfer of lead between the blood and bone tissues, and elimination of lead from the blood follows a first-order kinetic relationship.

While the adopted model is most certainly an oversimplification, model results will approximate those of other more complicated models involving additional tissue compartments for two reasons:

- While lead does mobilize from non-bone tissues following a decrease in environmental lead uptake, the effects are believed to be limited to a period of days or weeks due to the lower concentrations of lead amassed in these tissues, and
- While all lead elimination from the body does not occur via a direct pathway from the blood, the kinetic parameters used in the model properly include these other pathways (endogenous fecal and via other soft tissues) as if they were directly from the blood.

Based on the model illustrated in Figure F2-1, blood-lead concentrations (PbB) after intervention would follow the relationship illustrated in Figure F2-2. More specifically, immediately after intervention there would be an initial drop from the pre-intervention PbB level (PbB_{Pre}) to achieve an immediate post-intervention PbB level ($PbB_{ImmPost}$). $PbB_{ImmPost}$ represents the blood-lead concentration that can be supported by the amount of lead being transferred from the bone. After this initial drop, blood-lead concentrations would follow an exponential decline toward the long-term post-intervention PbB level ($PbB_{LongTerm}$). $PbB_{LongTerm}$ is the blood-lead level that can be supported by the post-intervention exposure level, with no additional lead from the bone. At any a particular length of time following the intervention, illustrated by the symbol "T" on the horizontal axis in Figure F2-2, a target post-intervention PbB level ($PbB_{Observed}$) will be observed. The original analysis using this model (Rust, et al., 1996) estimated the maximum length of time (T) the bone-lead stores would be capable of keeping the blood-lead concentration above the targeted observed level ($PbB_{Observed}$) for a given value of $PbB_{LongTerm}$. For the purposes of the sensitivity analysis for §403, the maximum long-term effectiveness is estimated instead. As the long-term percent decline reflects the post-intervention PbB that can be support by the post-intervention exposure level, it is assumed this decline is equal to the primary prevention effectiveness of the intervention.

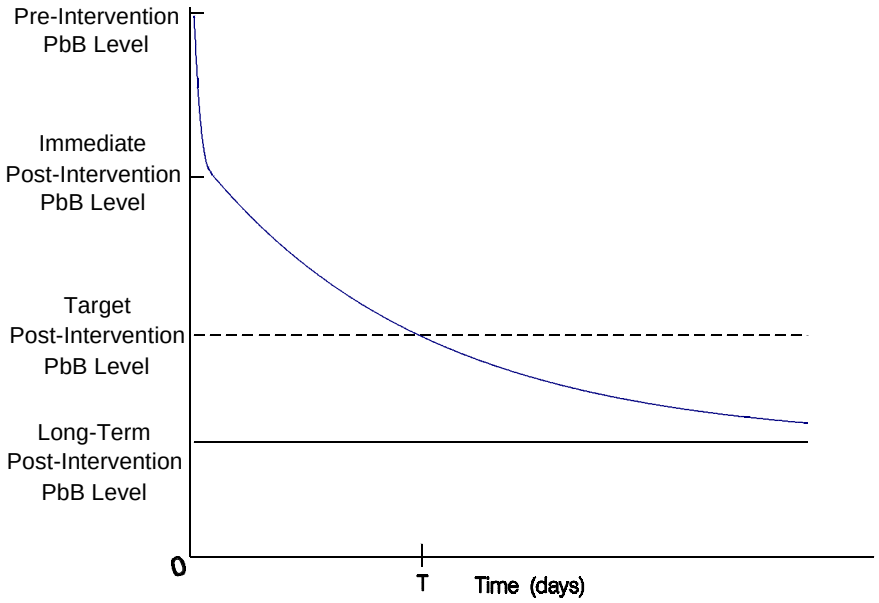


Figure F2-2. Blood-Lead Concentration Versus Time Following a Reduction in Lead Uptake.

The child's blood-lead concentration at t days post-intervention is given by the equation

$$PbB = PbB_{LongTerm} + (PbB_{ImmPost} - PbB_{LongTerm}) \cdot \exp(-t \cdot KBONEBL_{Net}) \quad (1)$$

where $KBONEBL_{Net}$ is the net rate of lead flow from bone to blood to elimination. This rate is a function of the blood-lead level following the initial drop ($PbB_{ImmPost}$) as well as other kinetic parameters (e.g., the lead mass ratio of bone to blood and the elimination rate of lead from the blood) which can be estimated from existing scientific literature (Rust, et al., 1996). As portrayed in Figure F2-2, the blood-lead concentration follows an exponential decline toward $PbB_{LongTerm}$. Setting PbB in Equation (1) equal to $PbB_{Observed}$ and solving for the long-term percent decline in blood-lead concentration ($R_{LongTerm}$) results in the following equation:

$$R_{LongTerm} = \frac{PbB_{LongTerm}}{PbB_{Pre}} = \frac{R_{Observed} - R_{ImmPost} \cdot \exp(-t \cdot KBONEBL_{Net})}{1 - \exp(-t \cdot KBONEBL_{Net})} \qquad (2)$$

where

$$R_{Observed} = \frac{PbB_{Observed}}{PbB_{Pre}} \quad \text{and} \quad R_{ImmPost} = \frac{PbB_{ImmPost}}{PbB_{Pre}} .$$

The maximum efficacy of an intervention, then, may be calculated given two parameters:

1. the observed percent decline ($R_{Observed}$) in an exposed child's blood-lead concentration following an intervention (i.e., the observed secondary prevention efficacy); and

2. the length of time (t) following the intervention when the decline was observed.

Note that this process estimates the maximum value of $R_{LongTerm}$ that might have yielded the inputted values of $PbB_{Observed}$ and t based on Equation (1). The specific value may lie between $R_{Observed}$ and $R_{LongTerm}$. The estimated primary prevention efficacy is a maximum in that $R_{ImmPost}$, and therefore $KBONEBL_{Net}$, cannot be estimated from available data (Rust, et al., 1996). It is necessary to estimate the maximum efficacy over a range of possible values for $R_{ImmPost}$.

Results of Modeling Bone-Lead Mobilization

To illustrate the efficacy of primary prevention, values of 25%, 50%, and 75% are considered for the observed secondary prevention efficacy and values of 6, 12, 18, and 24 months are considered for the lengths of time. Table F2-1 presents the maximum primary prevention efficacy for these scenarios for children 1 to 7 years of age. The standard error of the estimated efficacy—calculated by propagating, through the model, the standard errors of the underlying model parameters—is enclosed in parentheses.

As an example of the results in Table F2-1, note that if the observed effectiveness of a secondary intervention is assumed to be 25% (i.e., PbB decline to 75% percent of the pre-intervention level) at 6 months post-intervention for a 2 year old, then the implied effectiveness of primary prevention will be at most 47%. The scientific literature reports secondary prevention efficacy of approximately 25% declines in blood-lead concentration 12 months following dust

abatements, lead-based paint abatements, elevated soil lead abatements, and intensive educational efforts (USEPA, 1995b). Depending upon the age of the child benefitting from the intervention, the results in Table F2-1 would suggest these interventions would prompt primary prevention efficacy of between 30% and 59% (column: “Length of Time, 12 Months”; row: “Observed Efficacy of Secondary Prevention, 25%”).

Empty cells in Table F2-1 indicate that those scenarios cannot possibly occur based on Equation (1). For example, for a 7 year old, the impact of mobilized bone-lead stores would result in less than a 25% decline in blood-lead concentration at 6 months, even for a 100% effective intervention. Estimates of primary prevention efficacy under these “impossible” scenarios are not meaningful and are therefore not shown.

Consistent with the limited data available on bone-lead mobilization, the standard errors in Table F2-1 are quite large. By incorporating the 95% upper confidence bounds on the maximum primary prevention efficacy, the resulting bounded estimates are 1.2 to 1.9 times larger than the mean estimates reported in the table.

As described above, this analysis estimates the maximum efficacy of primary prevention interventions. Consideration was also given to obtaining the minimum efficacy. It was determined that the present model can provide a meaningful solution for the maximum case only, and that additional empirical data and extensive model enhancement are required to solve the minimum case. Only the maximum efficacy, therefore, is reported.

Table F2-1. Maximum Efficacy of Primary Prevention For Blood-Lead Levels (PbB) Observed at 25%, 50%, and 75% of Pre-Intervention Levels at 6, 12, 18, and 24 Months.

Observed Efficacy of Secondary Prevention[a]	Child's Age (years)	Length of Time[b] (months)			
		6	12	18	24
25%	1	0.39 (0.16)	0.30 (0.05)	0.28 (0.03)	0.27 (0.02)
	2	0.47 (0.18)	0.33 (0.08)	0.30 (0.04)	0.28 (0.03)
	3	0.56 (0.21)	0.36 (0.14)	0.31 (0.07)	0.29 (0.04)
	4	0.67 (0.25)	0.41 (0.19)	0.34 (0.10)	0.31 (0.06)
	5	0.79 (0.27)	0.47 (0.19)	0.37 (0.14)	0.33 (0.08)
	6	0.91 (0.32)	0.53 (0.21)	0.40 (0.19)	0.35 (0.12)
	7		0.59 (0.22)	0.44 (0.19)	0.37 (0.15)
50%	1	0.78 (0.32)	0.60 (0.09)	0.56 (0.05)	0.55 (0.04)
	2	0.94 (0.36)	0.65 (0.16)	0.59 (0.08)	0.56 (0.06)
	3		0.73 (0.27)	0.63 (0.13)	0.59 (0.08)
	4		0.83 (0.37)	0.68 (0.21)	0.62 (0.13)
	5		0.93 (0.38)	0.73 (0.29)	0.66 (0.17)
	6			0.81 (0.37)	0.70 (0.24)
	7			0.89 (0.37)	0.75 (0.31)
75%	1		0.90 (0.14)	0.84 (0.08)	0.82 (0.05)
	2		0.98 (0.25)	0.89 (0.13)	0.85 (0.09)
	3			0.94 (0.20)	0.88 (0.13)
	4				0.93 (0.19)
	5				0.98 (0.25)
	6				
	7				

Note: An empty cell means that the scenario is not possible according to model predictions.

[a] This is equivalent to the observed percent decline in an exposed child's blood-lead levels at a specified time point following the intervention.

[b] This is equivalent to the length of time following the intervention when the decline was observed.

APPENDIX G

Multi-Media Model (Empirical Model) for Use in the Section 403 Risk Assessment

APPENDIX G

MULTI-MEDIA MODEL (EMPIRICAL MODEL) FOR USE IN THE SECTION 403 RISK ASSESSMENT

EXECUTIVE SUMMARY

Purpose Of The Appendix

This appendix documents development and evaluation of an empirical regression model relating measures of lead in a residential environment to geometric mean children's blood-lead concentrations. The model is used as one tool in the Section 403 risk assessment to estimate blood-lead concentrations of children exposed to lead in paint, dust and soil as measured in the HUD National Survey. This model is also employed to evaluate various options for risk management for the Section 403 standards. In this analysis, EPA estimated a national distribution of blood-lead levels (and, ultimately, estimated health effects) before enactment of the Section 403 standards, and then employed models to relate environmental levels of lead to children's blood-lead levels to estimate a national distribution of blood-lead levels (and health effects) after enactment of specific 403 standards. Environmental measures of lead from the HUD National Survey are used as inputs to the empirical model to predict the national distribution of blood-lead concentrations. Therefore, the model development was constrained to variables in the HUD National Survey data set. The goal was to develop a model that could be used to give an approximation of expected blood-lead concentrations related to residential environmental lead based on a single source of data.

In this appendix the empirical model is presented and its prediction of a national distribution of blood-lead concentrations is compared to the results of Phase 2 of the Third National Health and Nutrition Examination Survey (NHANES III).

Model Development Issues

The choice and construction of variables, the mathematical form of the empirical model, assessment of goodness-of-fit and influential points, and the treatment of measurement error in predictor variables were all given consideration during the development of the empirical model.

One particular difficulty was that the empirical model was constructed using dust lead results collected from wipe sampling in the Rochester study, whereas dust lead results in the HUD National Survey were collected from blue nozzle vacuum sampling. Similarly, the empirical model was constructed using soil lead concentrations observed from drip-line sample locations in the Rochester study, whereas soil lead results in the HUD National Survey were based on an average concentration of lead in soil from drip-line, entryway and remote locations. A statistical method was developed to account for both systematic differences as well as differences in error structures between the sampling methods employed in the Rochester study and the HUD National Survey.

The Empirical Model

The form of the empirical model is:

$$\ln(PbB) = \beta_0 + \beta_1 \cdot \ln(PbF_{BN}) + \beta_2 \cdot \ln(PbW_{BN}) + \beta_3 \cdot \ln(PbS) + \beta_4 \cdot PbP + e$$

where PbB represents the blood-lead concentration, PbF_{BN} and PbW_{BN} correspond to dust-lead loading from interior floors and window sills respectively (on a Blue Nozzle Vacuum Scale), PbS represents soil-lead concentration, PbP represents paint/pica hazard, and e represents the residual error left unexplained by the model.

Results Of The Comparison With NHANES III

The predicted distribution of blood-lead concentrations for children aged 1-2 years obtained by applying the empirical model to the HUD National Survey Data was compared to Phase 2 of NHANES III. Results of this comparison indicate:

- The national geometric mean blood-lead concentration (pre-intervention) was properly calibrated to the geometric mean reported in NHANES III.

- The variability in the national distribution of blood-lead concentrations predicted by the empirical model using the HUD National Survey is approximately 1.71 (GSD), in contrast to a GSD of 2.09 for Phase 2 of NHANES III.

- The estimated proportions of blood-lead concentrations exceeding 10, 20 or 30 µg/dL using the empirical model predictions are much lower than the corresponding proportions estimated by NHANES III. For example, the percentage of children aged 1-2 years estimated to have blood-lead concentrations above 10 µg/dL using the empirical model was 1.54% in comparison to 5.75% estimated in Phase 2 of NHANES III.

Differences between the Rochester study population and the national population represent the primary limitation when using the empirical model based on data from the Rochester Study to predict a national distribution of blood-lead concentrations.

Use Of The Empirical Model

The empirical model is used in the Section 403 risk assessment and economic analyses to predict a distribution of childhood blood-lead concentrations based on measures of lead in paint, dust and soil at the child's primary residence. This information is used to evaluate various options for risk management for the proposed Section 403 Standards. In these analyses, the model is used to predict national distributions of children's blood-lead concentrations both before and after the rule is proposed. Estimates of environmental levels of lead before and after enactment of the Section 403 standards and after interventions resulting from the standards will be used as inputs to the model. The empirical model should only be used to predict a distribution of blood-lead

levels when environmental levels for all media are known or estimated. It is not intended as a general dose-response model, but rather as a predictive model developed specifically for use in the Section 403 Risk Assessment and specifically to predict blood-lead concentrations from estimates of environmental lead as measured in the HUD National Survey or as measured by a standard Section 402 risk assessment.

G1.0 INTRODUCTION

In order to better inform risk managers as they consider various options for the Section 403 standards, EPA estimated the range of risk reductions that are expected to result from a variety of potential standards. In order to do this, EPA estimated a national distribution of blood-lead levels (and, ultimately, potential health effects) before enactment of the Section 403 standards, and then relied on models relating environmental levels of lead to children's blood-lead levels to estimate a national distribution of blood-lead levels (and potential health effects) after enactment of specific 403 standards. The empirical model is used in the Section 403 risk assessment and economic analysis to predict a distribution of blood-lead concentrations related (jointly) to measures of lead in three media at the child's primary residence: paint, dust, and soil. Environmental measures of lead from the HUD National Survey were used as inputs to the empirical model to predict the national distribution of blood-lead concentrations. Therefore, the model development was constrained to variables in the HUD National Survey data set. Given time and budget constraints the goal for the empirical model development could not include construction of the best possible model based on multiple data sources. Rather, the goal was to develop a model that could be used to give an approximation of expected blood-lead concentrations related to residential environmental lead based on a single source of data. This model has not undergone formal validation and is based on only one data set. It is not intended as a general dose-response model, but rather as a predictive model developed specifically for use in the Section 403 Risk Assessment and specifically to predict blood-lead concentrations from estimates of environmental lead as measured in the HUD National Survey or as measured by a standard Section 402 risk assessment. The model was used to estimate the benefits of the 403 rule in the post-403 situation by estimating the reduction in children's blood lead concentrations resulting from application of various options for the 403 standards via risk assessments in residential housing.

In this appendix the empirical model is presented and its prediction of a national distribution of blood-lead concentrations is compared to the results of the NHANES III Survey as follows:

> A national distribution of housing and population characteristics was estimated using the HUD National Survey of environmental levels of lead in paint, dust, and soil in residential housing along with pertinent Census information. The Census information and the HUD National Survey measurements of environmental lead (after appropriate conversions) were used as inputs to the model to predict a national distribution of children's blood-lead levels before enactment of the Section 403 standards. This pre-rulemaking distribution was compared to the national distribution of children's blood-lead concentrations estimated by the NHANES III survey to assess the adequacy of the model and its applicability on a national level.

The empirical model is also used to predict the national distribution of children's blood-lead levels after enactment of the Section 403 standards. Estimates of environmental levels of lead after the conduct of interventions performed in response to various options for the Section

403 standards are used as inputs to the model. Comparison of the pre- and post-rulemaking distributions allow estimation of the benefits associated with the rulemaking.

The empirical model is not intended to be used to estimate the effect of a single media on blood-lead levels. The model should only be used to predict a distribution of blood-lead levels when environmental levels for all media are known or estimated. Individual parameter estimates should not be interpreted in isolation.

The choice and construction of variables, the mathematical form of the empirical model, assessment of goodness of fit and influential points, and the treatment of measurement error in predictor variables were all given consideration during the development of the empirical model, and are described in detail in this document.

G2.0 DESCRIPTION OF DATA

The purpose of the statistical analysis was to provide a predictive model which relates childhood blood-lead concentration to measures of lead exposure from paint, dust and soil. Variables which represent lead exposure in environmental media were based on data that was available in both the HUD National Survey, and in the Rochester Lead-in-Dust study. Data from the HUD National Survey and from NHANES III are based on surveys that were designed to be nationally representative of the housing stock and the population of children, respectively. The HUD National Survey was a survey of pre-1980 housing that was adjusted using data from the 1993 American Housing Survey to represent the 1997 housing stock as described in the Section 403 Risk Assessment document. The Rochester Study was based on a targeted sample limited to a single geographic area as were other candidate epidemiological (epi) studies. It is unclear as to whether inferences drawn from any particular epi-study can be generalized to the national population of children and/or housing. Following is a brief discussion of each individual source of data, as well as a rationale and description of the variables that were included in the statistical analyses.

G2.1 SOURCES OF DATA

G2.1.1 Rochester Lead-in-Dust Study

The Rochester Lead-in-Dust Study is a cross sectional study which recruited 205 children from live births at three local hospitals using a stratified sampling scheme. The sampling scheme was designed to recruit a high proportion of low income families living in older (pre 1940) housing. Blood-lead and hand-lead sample collection from recruited children occurred between August 31 and November 20, 1993. A detailed questionnaire was also completed at the time of blood sample collection. Environmental assessment of the primary residence of each recruited child was generally completed within three weeks of the date of blood sample collection, and included samples of dust from floors, window sills and wells, samples of soil from the dripline adjacent to the foundation and the child's play area, and measurements of painted interior and exterior surfaces (condition of paint and XRF paint lead loading).

G2.1.2 HUD National Survey

The HUD National Survey collected environmental samples of paint, dust, and soil from 284 private homes between 1989 and 1990. The objective of the study was to obtain data for estimating the prevalence of lead-based paint and lead-contaminated dust and soil in the nation. The presence or absence of children with elevated blood-lead was not part of the sampling design. One floor-dust sample was collected from each of three rooms, and one window sill and window well sample was collected from each of two rooms using a blue nozzle vacuum sampler. Three soil samples were collected from the dripline, entryway and remote locations. Paint sampling included XRF measures of paint-lead loading and condition of paint from generally two interior rooms and one side on the exterior of each residential unit.

In the HUD National Survey, each unit was assigned a sampling weight equal to the number of pre-1980 privately-owned, occupied units in the national housing stock that were represented by the given unit in the survey. The total of all 284 sampling weights equaled the number of pre-1980 privately-owned, occupied units in the national housing stock at the time of the survey. Sampling weights in the National Survey were determined according to four demographic variables associated with the units:

- Age category of unit
- Number of units in the building
- Census region
- Presence of a child under age 7 years

Since EPA's Risk Assessment uses 1997 as a base year for Section 403 activities, it was desirable to use the environmental-lead levels from the National Survey to characterize environmental-lead levels in the 1997 national housing stock. Therefore the sampling weights of National Survey units were revised to represent the 1997 occupied housing stock. The revised weights indicate the number of units in the 1997 national housing stock that are associated with the given National Survey unit, and therefore, with its distribution of environmental-lead levels.

G2.1.3 National Health and Nutritional Educational Survey (NHANES) III

The Third National Health and Nutrition Examination Survey (NHANES III), conducted from 1988 to 1994, was the seventh in a series of national examination studies conducted by CDC's National Center for Health Statistics (NCHS) to trace the health and nutritional status of the non-institutionalized, civilian U. S. population. The target population for NHANES III included the civilian non-institutionalized population 2 months of age and older.

To provide for a nationally representative sample and sufficient precision in characterizing key subgroups, a complex survey design was employed in NHANES III. Approximately 40,000 persons were sampled in NHANES III, including approximately 3,000 children aged 1 to 2 years. Although estimates of national population health and nutrition parameters were the primary objectives of the survey, suitably precise estimates for certain age and race groups were obtained through over sampling. As a result, the NHANES III provides a solid basis for obtaining national estimates of the distribution of childhood blood-lead concentrations. Details on the study design and how the survey was conducted are available from CDC, 1992 and CDC, 1994.

G2.1.4 Other Candidate Epi Studies Considered

There are various other epi studies that were potential data sources on which to base the empirical model. Given time and budget constraints the goal for the empirical model development could not include construction of the best possible model based on multiple data sources. Rather the goal was to develop a model that could be used to give an approximation of expected blood-lead concentrations related to residential lead based on a single source of data. The Rochester Study was chosen because of the following advantages:

1. All media, locations, and surfaces that are being considered for Section 403 standards were measured for lead in the Rochester Study.

2. The Rochester Study includes dust-lead loadings from wipe sampling and the Section 403 dust standard is expected to be based on dust-lead loading from wipe sampling.

3. The selection of homes and children in the Rochester Study, although targeted, was more random and more representative of a general population than is the case with most recent epidemiological studies of lead exposure in non-smelter communities.

4. The Rochester Study is recent.

The primary limitation associated with the Rochester Study is concern over the degree to which the Rochester Study may be considered representative of the nation as a whole. The limitations of the Rochester Study are discussed in more detail in Section G8.

Other data sets considered for use in constructing the empirical model included:

1. Pre-intervention data from the Baltimore Repair and Maintenance (R&M) Study. The R&M Study is a prospective longitudinal study which was designed to investigate the potential health and environmental benefits associated with performing R&M interventions on urban housing with lead-paint hazards. The pre-intervention sample included 115 children living in 87 homes. Samples of blood were collected from each participating child, and samples of dust, soil and water were collected from each house during the pre-intervention campaign. Due to the fact that the housing stock in this study consisted primarily of Baltimore City rowhouses, only 42 children living in 29 homes had soil samples. The absence of measures of lead in soil would have limited the use of this data in the development of an empirical model focused on all three media: paint, dust and soil.

2. Pre-intervention data from the Boston Soil Lead Abatement Demonstration Project. The Boston 3-City Study recruited 152 children living in 101 houses from four different urban neighborhoods during the pre-intervention campaign. The main restrictions for recruitment into the study were that the children had to be under the age of 5 and have an initial blood-lead concentration between 7 and 24 µg/dL. For each household recruited into the study, a detailed environmental assessment was conducted concurrently with the blood-sampling. This environmental assessment included the collection of samples from paint, dust, soil and water. All dust samples from the Boston 3-City Study were collected using the Sirchee-Spitler Method. This method entails the use of a modified Black & Decker Dustbuster vacuum, and its properties with respect to other sampling methods are not well understood at the current time. Collection of a handwipe sample from each participating child and the completion of a questionnaire was also conducted with each blood sample.

The restricted range of blood-lead concentrations recruited into this study was likely to have a large impact on parameter estimates of the relationships under investigation, and therefore, this source of data was not considered optimal for use in developing the empirical model.

3. Pre-intervention data from the Baltimore Soil Lead Abatement Demonstration Project. The Baltimore 3-City Study recruited 402 children living in 204 houses from two different urban neighborhoods during three rounds of pre-intervention sampling. There were no restrictions on the blood-lead concentration of children recruited into the study, however children had to be under the age of seven. For each household recruited into the study, a detailed environmental assessment was conducted once during the pre-intervention campaign. This environmental assessment included the collection of samples from dust, soil, exterior paint, and water. The Baltimore 3-City Study did not include samples of lead in paint or dust from window sills or window wells. Samples of interior paint were collected after the soil abatement intervention took place. In addition, all dust samples from the Baltimore 3-City Study were collected using the Sirchee-Spitler Method, and its properties with respect to other sampling methods are not well understood at the current time. Therefore, this source of data was not considered optimal for use in developing the empirical model.

4. Pre-intervention data from the Cincinnati Soil Lead Abatement Demonstration Project. The Cincinnati 3-City Study included 201 children living in 129 houses from six different urban neighborhoods in the first (pre-intervention) phase of the study. The households recruited into the study were mostly single family residential units within multi-unit apartment buildings. It was believed that lead-based paint was removed from participating residential units in the early 1970's as part of a housing rehabilitation project. The pre-intervention environmental assessment consisted of the collection of interior and exterior dust and paint from each participating residential unit, and samples of soil from neighborhood recreation areas such as parks and playgrounds. Dust samples were collected using the DVM sampling method. Soil abatement was performed on a neighborhood scale, in parks, play areas, and other common grounds. Exterior dust was also removed from the neighborhood streets, alleys, and sidewalks as part of the intervention. Since soil samples could not be related to individual residences, this source of data was not considered optimal for use in developing the empirical model.

5. Data from the Cincinnati Longitudinal Study. The Cincinnati Longitudinal Study is a prospective study which followed a cohort of several hundred children from birth to five years of age. It was designed to assess the impact of urban lead exposure on children's blood-lead concentrations. Once a year, blood-lead and hand lead samples were collected from each participating child. Progress in social, behavioral and cognitive development for each child was also measured over the course of the study. Environmental samples which included interior surface dust, XRF paint and exterior surface scrapings were collected from the residences of each participating child at approximately the same time as blood sample collection. There was also a qualitative

housing evaluation that was conducted for each residence included in the study. The Cincinnati Longitudinal Study provides data on the relationship between blood-lead and environmental lead over time. Although it is uncertain as to whether the exterior surface scrapings are representative of exterior dust or soil (or both), it appeared as though the Cincinnati Longitudinal Study was a good potential source of data for the empirical model; however these data have not yet been publicly released by the University of Cincinnati.

6. Data from the HUD Lead-Based Paint Hazard Control Grant Program in Private Housing (HUD Grantee data). HUD has provided grants to states and units of general local government (Grantees) for environmental interventions in privately owned low- and moderate-income housing. HUD requires Grantees to conduct dust-wipe testing and blood testing prior to environmental intervention. Paint and soil sampling are optional. Data from this program was not available for analysis at the time of preparation of the empirical model.

G2.2 VARIABLES UNDER CONSIDERATION

Following is a rationale and description of the variables that were most closely examined for inclusion in the empirical model. These variables represent a subset of all the variables originally considered. They were selected based on several properties, including strength of association with blood-lead concentration in bivariate models, predictive power when included into a model with competing sources of lead exposure, interpretation, ability to construct the variable across different sources of data, and applicability to data collected by a standard Section 402 risk assessment.

The criteria used for the selection of variables in the empirical model emphasized use of measures of environmental lead and other factors observed in both the Rochester Lead-in-Dust Study and the HUD National Survey. Variables whose definition provided a convenient translation when applied to the National Survey, whose predictive power in Rochester were high, and whose spread in the National Survey populations covered a wide enough range of values, were used in the empirical model.

The first group of variables are subject specific, constructed from measurements on each child recruited into the empirical studies. The second group of variables are property specific, representing observations from the primary residences of each of the subjects. Because the Rochester Study included only one child per household, all of the variables measured in this statistical analysis can be organized using an identifier for household, represented by the subscript, i, throughout this document.

G2.2.1 Subject Specific Variables

Table G-1 gives descriptions of the subject-specific variables: blood-lead concentration, age, pica and race.

Table G-1. Subject-Specific Variable Descriptions

Variable	Description
Blood-Lead	Blood lead concentration on a venous sample is reported in units of micrograms of lead per deciliter (µg/dL). Because the distribution of blood-lead concentration is usually skewed, a natural log transformation was applied to blood lead concentration for use as a response variable in the statistical models. The natural log transformation helps the distribution of observed blood-lead levels meet normality assumptions required by the statistical models. $LPbB_i$ = Natural log of the blood lead concentration measured from the ith child.
Pica	It has been hypothesized that sources of lead exposure in environmental media influence blood-lead concentration as a function of the hand-to-mouth activity or mouthing behavior of the child. A child who exhibits "strong" mouthing behavior or pica may be at higher risk for attaining an elevated blood-lead concentration. The following two questions were included in the Rochester Lead-in-Dust study as part of the questionnaire, and were designed to measure mouthing behavior or pica tendencies in children: (1) How often does the child put paint chips in his/her mouth?, and (2) How often does the child put dirt or sand into his/her mouth? The following choices were given as a possible response to these questions. 0 Never 3 Often 1 Rarely 4 Always 2 Sometimes The following Pica variables were constructed based on the parental/guardian responses to the above two questions: $Paint\ Pica_i$ = Tendency of the ith child to put paint chips in the mouth (on a scale of 0 to 4). $Soil\ Pica_i$ = Tendency of the ith child to put dirt or sand in the mouth (on a scale of 0 to 4).
Age	Age has been documented as having a nonlinear effect on blood lead concentration when children are young (CDC, 1991). Therefore the age of each subject (in years) measured at the time of blood sampling was considered as a potential covariate in the statistical analysis. Age_i = Age (continuous measure in years) of the ith child.
Race	It is quite possible that there are biological, cultural and/or behavioral differences among children recruited into the Rochester study that cannot be explained by any of the other measured variables barring race. Indicator variables representing race were therefore explored as covariates for the statistical analyses: $White_i$ = 1 if the ith child is Caucasian. = 0 Otherwise $Black_i$ = 1 if the ith child is of African American descent. = 0 Otherwise $Other_i$ = 1 if the ith child is not Caucasian or not African American. = 0 Otherwise

G2.2.2 Property Specific Variables

The property specific variables that were investigated in this statistical analysis correspond to measures of lead exposure from paint, dust and soil. There are many different ways of constructing lead exposure variables from the various different samples that were collected from

each environmental media. The variables discussed below represent one way of characterizing lead levels in environmental media.

Table G-2. Property-Specific (Dust and Soil) Variable Descriptions

Exposure	Description
Paint (75th Percentile)	Interior and exterior paint lead loading was measured on multiple different painted surfaces within each residential unit using portable XRF instruments. Usually the condition of the paint was also measured for each painted surface that was sampled. Several variables were constructed using a combination of observed paint lead loadings and condition of the paint from both the interior and exterior of each residential unit. Two variables were chosen for the statistical analyses, which represent the presence and severity of deteriorated interior and exterior lead-based paint. The following formula describes the construction of the paint-lead variables, and was applied separately for interior and exterior paint samples within each residential unit: Let XRF_{ij} represent the observed paint lead loading (mg/cm^2) from the jth component within the ith residential unit, if the XRF value was greater than or equal to 1 mg/cm^2. An observed XRF paint-lead loading greater than or equal to one is considered lead-based paint. If the observed paint lead loading was less than 1 mg/cm^2, XRF_{ij} is equal to zero. Condition of the paint is characterized as Good whenever less than 5% of the surface is deteriorated; Fair whenever 5% to 15% of the surface is deteriorated; and Poor whenever more than 15% of the surface is deteriorated. By combining categories, let $Cond_{ij}$ represent the condition of the paint on the jth component within the ith residential unit; $Cond_{ij}$ is equal to one if the surface was rated Fair or Poor, and is equal to zero if it was rated Good. Then we have a measure of deteriorated LBP, which is given by $DETLBP_{ij} = XRF_{ij} \cdot Cond_{ij}$ $Paint_i$ is defined as the 75th percentile of the j observed levels of $DETLBP_{ij}$. It is a variable which represents the presence and severity of deteriorated lead-based paint within a residential unit. Residential units in which less than 25% of the sampled painted surfaces had deteriorated lead-based paint result in a $DETLBP_{ij}$ value that is equal to zero. Residential units with 25% or more of the sampled painted surfaces having deteriorated lead-based paint result in $DETLBP_{ij}$ values that are greater than or equal to one. Int_pnt_i = $Paint_i$ based on interior painted surfaces. Ext_pnt_i = $Paint_i$ based on exterior painted surfaces.
Paint/Pica Hazard	An additional paint variable combined paint condition, lead-based paint and pica. An indicator variable which was nonzero whenever each of the following conditions existed in a residential unit: presence of deteriorated or damaged interior paint in the household; and presence of interior lead-based paint in the household; and presence of a child with paint pica in the household. The paint variable had values of: 0 No LBP (XRF reading < 1), or condition is Good, or child does not exhibit paint pica; 1 LBP (XRF reading ≥ 1), condition is Fair or Poor, and child exhibits paint pica rarely; 2 LBP (XRF reading ≥ 1), condition is Fair or Poor, and child exhibits paint pica at least sometimes. In the Rochester Study, a child's tendency towards paint pica was characterized as: 0 = Never, 1 = Barely, 2 = Sometimes, 3 = Often and 4 = Always. Because of limited sample size in each category, Paint pica was collapsed for this modeling to have values: 0 = No paint pica, 1 = Child exhibits paint pica rarely, and 2 = Child exhibits paint pica at least sometimes. A value of 1.5 was chosen as the input value for those children exhibiting pica at least rarely in applying the empirical model to the HUD National Survey. The average value of this pica variable for children who exhibited any pica in the Rochester Study was 1.25

Table G-2. Property-Specific (Dust and Soil) Variable Descriptions (Continued)

Exposure	Description
Floor Dust Combined With Proportion of Carpeted/ Uncarpeted Surfaces	There were residential units in which all floor surfaces that were sampled were either carpeted or uncarpeted, resulting in missing values for the variables $Floor\_C_i$ or $Floor\_U_i$. A second set of floor-dust exposure variables were therefore pursued in an effort to recapture residential units with missing values. Let PC_i. represent the proportion of floor dust samples collected from carpeted surfaces within the ith house: PC_i = [Number of carpeted floor surfaces]$_i$ / [Total number of floor surfaces sampled]$_i$ Then $Carp\_flr_i = Floor\_C_i * PC_i$, and $Bare\_flr_i = Floor\_U_i * (1-PC_i)$ where $Carp\_flr_i$ represents the area weighted arithmetic average dust-lead loading from carpeted floors multiplied by the proportion of floor dust samples that were collected from carpeted surfaces in the ith residential unit. Note that $Carp\_flr_i$ is equal to zero for residential units that had no carpeted surfaces sampled. $Bare\_flr_i$ represents the area weighted arithmetic average dust-lead loading from uncarpeted floors multiplied by the proportion of floor dust samples that were collected from uncarpeted surfaces in the ith residential unit. Note that $Bare\_flr_i$ is equal to zero for residential units that had no uncarpeted surfaces sampled.
Dust (Window Trough, Window Sill and Floor)	Samples of interior household dust were collected from floors, window sills and window wells from residential units in the Rochester Study. Dust samples were collected using both wipe and vacuum samples, thus measures of dust-lead loading were available for all dust samples, and measures of dust-lead concentration are available for those dust samples that were collected using vacuum samples. Variables were constructed which represent the area weighted arithmetic average dust-lead loading and the mass weighted arithmetic average dust-lead concentration for each component type tested within each residential unit. Due to a lack of understanding of potential differences between the exposure mechanism between carpeted and uncarpeted surfaces, floor dust samples collected from carpeted and uncarpeted surfaces were treated as separate component types in the construction of variables. An initial assessment comparing dust-lead loading variables to dust-lead concentration variables (for samples collected using vacuum sampling) in the Rochester Lead-in-Dust Study demonstrated that the lead-loading variables were consistently stronger predictors of blood-lead concentrations. In addition, it is expected that dust standards will be specified in terms of dust-lead loading from wipe samples. Therefore, the following measures of wipe dust-lead loading were considered as potential variables in the predictive model: $Floor\_A_i$ represents the area weighted arithmetic average dust-lead loading from all surface (carpeted or uncarpeted) floors in the ith residential unit. $Floor\_C_i$ represents the area weighted arithmetic average dust-lead loading from carpeted floors in the ith residential unit. $Floor\_U_i$ represents the area weighted arithmetic average dust-lead loading from uncarpeted floors in the ith residential unit. W_Sill_i represents the area weighted arithmetic average dust-lead loading from window sills in the ith residential unit. W_Well_i represents the area weighted arithmetic average dust-lead loading from window wells in the ith residential.

Table G-2. Property-Specific (Dust and Soil) Variable Descriptions (Continued)

Exposure	Description
Soil	Composite samples of soil were collected using a coring tool from several different locations within the yard of each residential unit. In the Rochester Lead-in-Dust Study, the laboratory analysis of the composite soil samples resulted in measures of soil-lead concentration (µg/g) for a fine soil fraction and a coarse soil fraction. An initial assessment of the soil-lead data from the Rochester Lead-in-Dust Study Data showed no statistically significant difference in predictive power between the fine and coarse soil fractions. Soil samples usually undergo some degree of sieving (note the HUD Guidelines protocol for soil sampling, Appendix 13.3, page App 13.3-3). Historically, the fine soil fraction has been used as a predictor variable in lead exposure studies, because it was thought that the fine-soil fraction is more bioavailable to children. We therefore considered only the fine-soil fraction in the statistical analyses. The following soil-lead exposure variables were considered as potential predictor variables in the statistical models: $Drip\_Soil_i$ represents the observed lead concentration in a composite soil sample collected from the dripline (adjacent to the foundation) of the ith home. $Play\_Soil_i$ represents the observed lead concentration in a composite soil sample collected from the play area of the ith home. Note that $Play\_Soil_i$ could be considered a subject specific variable.

G3.0 FORMS OF THE STATISTICAL MODELS

This section contains a discussion of the different forms of mathematical models considered for characterizing the relationship between blood-lead and measures of lead exposure that were considered as part of the modeling effort. The following five mathematical model forms were investigated for the development of a multi-exposure predictive model for childhood blood-lead concentrations. Each model is individually discussed in terms of statistical assumptions, biological/physical assumptions, and mathematical ease of use. Although biological/physical plausibility is an important issue, the objective of the empirical Model was to predict a rational distribution of blood-lead concentrations. Thus, the primary basis for choosing a model was based on predictive ability. It should be noted that there is currently no definitive model accepted by the scientific community for the relationship between childhood blood-lead and environmental-lead. The final form of the empirical model is presented in Section G6.

G3.1 LOG-LINEAR MODEL

The log-linear model expresses natural-log transformed blood-lead concentration as a linear combination of natural-log transformed exposure variables and select covariates. A multimedia exposure log-linear model for blood-lead concentrations (in generic form) would appear as follows:

$$\ln(PbB_i) = \beta_0 + \beta_1 \cdot \ln(Dust_i) + \beta_2 \cdot \ln(Soil_i) + \beta_3 \cdot \ln(Paint_i) + \gamma \cdot Covariate_i + e_i$$

where e_i (the residual error) is assumed to follow a normal distribution with mean zero and variance σ^2_{Error}.

One main advantage of the log-linear model is its mathematical convenience. The log-linear model is easily fitted using standard linear regression methods (although in the development of a multiple-exposure model it may be necessary to fit the log-linear model using a numerical approximation method while constraining parameter estimates for exposure variables to positive values; i.e. β_1, β_2, and $\beta_3 \geq 0$). Another mathematical convenience of the log-linear model is the fact that calculation of tolerance intervals and exceedance proportions, and adjusting for the effects of measurement error in predictor variables is relatively straight-forward.

With respect to biological/physical assumptions, the log-linear model when translated back into the original scale of observed blood-lead concentrations, results in a multiplicative relationship for environmental-lead:

$$PbB_i = \exp(\beta_0) \cdot Dust_i^{\beta_1} \cdot Soil_i^{\beta_2} \cdot Paint_i^{\beta_3} \cdot Covariate_i^{\gamma} \cdot \exp(e_i)$$

Thus, the effect of dust-lead on blood-lead is dependant on the combined effects of all of the other variables included in the model. Furthermore, the difference in predicted blood-lead

concentration for children exposed to dust-lead loadings of 5 and 50 $\mu g/ft^2$ is the same as the difference in predicted blood-lead concentration for children exposed to dust-lead loadings of 500 and 5000 $\mu g/ft^2$. Although the multiplicative interpretation of the log-linear model is not considered biologically/physically plausible, it often fits the data better than statistical models with a more plausible, biological/physical basis for data with low to moderately exposed children (Rust, et al., 1996).

G3.2 LOG-ADDITIVE MODEL

Whereas the log-linear model when translated back to the original scale of measurement results in an assumed multiplicative relationship, the log-additive model results in an assumption of additivity among the exposure variables. The log-additive model expresses natural-log transformed blood-lead concentration as the natural-log of a linear combination of exposure variables and select covariates. A multimedia exposure log-additive model for blood-lead concentrations (in generic form) would appear as follows:

$$\ln(PbB_i) = \ln(\beta_0 + \beta_1 \cdot Dust_i + \beta_2 \cdot Soil_i + \beta_3 \cdot Paint_i + \gamma \cdot Covariate_i) + e_i$$

where e_i (the residual error) is assumed to follow a normal distribution with mean zero and variance σ^2_{Error}.

Since the response variable in the log-additive model is expressed as a non-linear function of the exposure variables, it must be fitted using a non-linear regression algorithm. Thus, the mathematical conveniences of the log-linear model do not apply to the log-additive model.

With respect to biological/physical assumptions, the log-linear model when translated back into the original scale of observed blood-lead concentrations, results in an additive relationship for environmental-lead:

$$PbB_i = (\beta_0 + \beta_1 \cdot Dust_i + \beta_2 \cdot Soil_i + \beta_3 \cdot Paint_i + \gamma \cdot Covariate_i) \cdot \exp(e_i)$$

Thus, the effect of each measure of environmental lead on blood-lead is not dependant on the combined effects of all of the other variables that were included in the model. The model is attractive in that it is reasonable and biologically plausible that the relationship between blood-lead and environmental lead would be additive at low levels of environmental exposure. However, there is also evidence that saturation of the effect of environmental lead on blood-lead concentration occurs at higher levels of lead exposure, in which case additivity may no longer hold.

G3.3 ALTERNATE LOG-ADDITIVE MODEL

Although the additive interpretation of the log-additive model is more biologically plausible than the multiplicative interpretation of the log-linear model, the tendency of the log-additive model to over predict blood-lead at higher levels of environmental lead exposure may present a problem. One method for solving the problem is to use mathematically transformed measures of environmental-lead (such as the natural-log transformation) in the log-additive model. This "Alternate Log-Additive Model" would preserve the additivity property associated with the log-additive model, while also accounting for a saturation of the effect of environmental lead on blood-lead concentration at higher levels of lead exposure. A multimedia exposure version of an alternate log-additive model for blood-lead concentrations (in generic form) would appear as follows:

$$\ln(PbB_i) = \ln[\beta_0 + \beta_1 \cdot \ln(Dust_i) + \beta_2 \cdot \ln(Soil_i) + \beta_3 \cdot \ln(Paint_i) + \gamma \cdot Covariate_i] + e_i$$

where e_i (the residual error) is assumed to follow a normal distribution with mean zero and variance σ^2_{Error}.

The alternate log-additive model must also be fitted using non-linear regression, and therefore the alternate log-additive model does not have the same mathematical conveniences that are associated with the log-linear model. When using the alternate log-additive model, particular attention should be paid to the mathematical transformation that is applied to the environmental lead exposure variables. A transformation that is too strong may result in a model in which the effect of saturation at high environmental-lead levels is over-predicted, resulting in a model which under-predicts blood lead.

G3.4 ACTIVE/PASSIVE UPTAKE MODEL

Another method of adjusting the log-additive model to compensate for saturation of the response at high levels of environmental lead is to parameterize the saturation effect itself. The following "Active/Passive Uptake" Model demonstrates one method for parameterizing the saturation effect:

Let $Exposure_i$ represent a linear combination of the exposure variables (on the original scale) similar to the linear combination that appears inside the natural-log function in the log-additive model;

$$Exposure_i = \beta_0 + \beta_1 \cdot Dust_i + \beta_2 \cdot Soil_i + \beta_3 \cdot Paint_i + \gamma \cdot Covariate_i$$

The Active/Passive Uptake Model is then expressed as:

$$\ln(PbB_i) = \ln(Exposure_i) + \ln\left(F_{Passive} + \frac{1 - F_{Passive}}{1 + \frac{Exposure_i}{\theta}} \right) + e_i$$

where $0 \leq F_{Passive} \leq 1$ and $0 < \theta$

Figure G-1 provides a plot of blood-lead concentration as a function of $Exposure_i$, assuming that θ=10 µg/dL and that $F_{Passive}$ takes on values of 0, 0.05, 0.1, 0.5 and 1. The plot shows that when $F_{Passive}$ is equal to zero, θ =10 µg/dL provides an asymptote for the maximum blood-lead concentration that is predicted as a function of $Exposure_{i.}$ In the Active/Passive Uptake Model, $F_{Passive}$ represents the portion of $Exposure_i$ which has a linear effect on blood-lead concentration beyond the saturation point of $\theta(1- F_{Passive})$. When $F_{passive}$ equals 1, the Active/Passive Uptake Model is identical to the log-additive model, and therefore does not compensate for saturation of the response at high levels of exposure.

Advantages include biological/physical plausibility, goodness of fit relative to other candidate models (as is seen in the tables of Section G13) and the fact that this model is similar in nature to the relationship modeled within the IEUBK model. Disadvantages include the fact that this model may overparameterize the relationship between blood-lead and environmental lead in these data. Also, the active/passive uptake model does not have the same mathematical conveniences associated with the log-linear model.

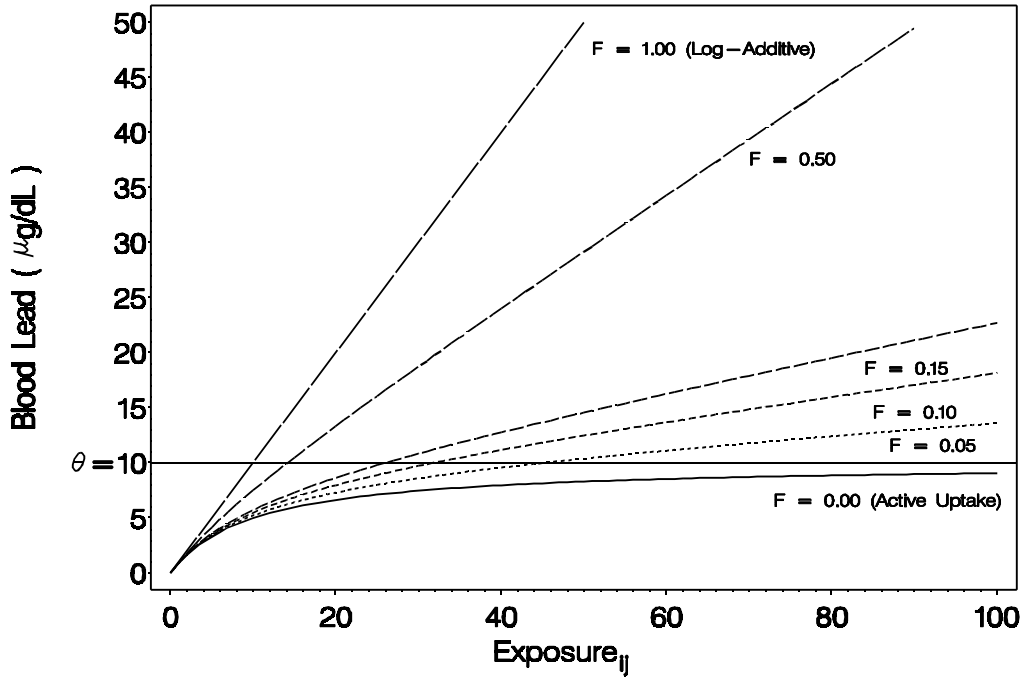


Figure G-1. Plot of Blood-Lead Concentration as a Function of Lead Exposure Using the Active/Passive Uptake Model with $\theta = 10$ µg/dL and $F_{Passive}$ Ranging from Zero to One.

G3.5 ACTIVE UPTAKE MODEL

The Active Uptake Model is simply a reduced form of the Active/Passive Uptake model in which the parameter $F_{Passive}$ is held fixed at zero. This model includes properties similar to the Active/Passive Uptake model, and may in some cases provide more interpretable parameter estimates for situations in which the Active/Passive Uptake Model is overparameterized.

G4.0 MEASUREMENT ERROR

The fact that the lead predictor variables for paint, dust and soil are subject to measurement error raises issues about the need to account for this measurement error in the model building process. In addition, the fact that different sampling methods were used in the Rochester Study and the HUD National Survey also raises issues about similar adjustments for different sampling methods when applying the empirical model to the HUD National Survey Data. Choosing an appropriate statistical methodology for adjustment is dependant on several factors including use of the model, interpretation of the predictor variables, the definition of the components of measurement error in the predictor variables, and the mathematical form of the model relating blood-lead to environmental lead.

Sections G4.1, G4.2, and G4.3 of this appendix discuss, respectively, the questions of:

1. what is being measured (and modeled) in the empirical model;
2. what adjustment for the effects of measurement error in predictor variables or differences in sampling methods is appropriate (with respect to Section 403 rulemaking activities);
3. the definition and characterization of measurement error associated with dust predictor variable.

G4.1 WHAT IS BEING MEASURED (AND MODELED)

The purpose of the empirical model is to assess the changes in the distribution of blood-lead levels of children one to two years old that are likely to result from the application of the 403 Rule standards. The vehicle for application of the 403 Rule standards is a risk assessment conducted in accordance with the work practice standards in the 402 Rule and following the detailed approach for risk assessments in the 1995 HUD Guidelines. Accordingly, the multi-media model defined in this document seeks to establish a relationship between children's blood-lead levels and environmental-lead levels as would be measured in a risk assessment. Environmental and blood-lead data from the Rochester Lead-in-Dust Study provided the means to develop the multi-media model. The relationship between blood-lead and environmental-lead observed in the Rochester Study was to be applied to environmental inputs from the HUD National Survey (with weights adjusted to represent housing in 1997). In most cases, environmental variables included in the multi-media model based on the Rochester data were constructed similarly using environmental-lead levels observed in the HUD National Survey. However, dust and soil measures were sufficiently different between these two studies, and a statistical adjustment procedure had to be developed to allow dust-lead and soil-lead measures from the HUD National Survey to be properly used as inputs to the model. This adjusted relationship between blood-lead and environmental-lead as observed in the HUD National Survey results in what this document refers to as the empirical model. For development of 403 Rule standards, the empirical model will be used to assess different options for the standards, and the resulting changes in the children's blood-lead distribution will be assessed to estimate the benefits

of the various options. Proper development of the empirical model requires attention to and balancing of three key features: 1) how environmental lead measurements are made in a risk assessment, 2) how environmental measurements were made in the Rochester Study and in the HUD National Survey, and 3) in the Rochester Study, what environmental measures are strong predictors of children's blood-lead levels.

The lead exposure variables used in the statistical model(s) were constructed from measured levels of lead in various different samples of paint, dust, or soil from the primary residence of each subject child. Protocols for environmental sampling were used in each study to assure that the measures of lead in environmental media were consistent across the various different houses recruited into that particular study. These protocols required detailed sampling in an effort to characterize the levels of lead in paint, dust, and/or soil at the time of sampling. The collection of environmental samples from a child's primary residence usually occurred within a few weeks of the collection of that child's blood-sample.

In the variable selection phase of the statistical analysis, various different ways of combining the lead-loading (or lead concentration) of same media samples within a residence into a lead exposure variable were investigated in terms of (1) their association with blood-lead in bivariate model(s), (2) their association with blood-lead in multimedia exposure model(s), and (3) ease of interpretation. In each case, the resulting variable was designed to characterize the child's exposure to lead in paint, dust, or soil ***from the primary residence***.

Although a child's blood-lead concentration is a product of cumulative exposure to lead, most of the available data from the lead exposure studies only provide information on the lead levels in environmental media at one point in time. Thus, the lead exposure variables that were constructed for use in the statistical models represent an estimate of the child's exposure to lead from paint, dust or soil from the primary residence at the time of sampling. The exposure variables (environmental lead) characterize current exposure to lead, rather than cumulative exposure to lead, whereas the response variable (blood-lead) is a measure of cumulative exposure. These exposure variables, including dust-wipe lead loadings, are similar to the measures that would be collected in a standard Section 402 risk assessment.

Therefore, the empirical model provides an estimate of the relationship between childhood blood-lead concentrations (indicative of a child's cumulative exposure to lead) and sampled measurements of lead from paint, dust, or soil ***from the primary residence at the time of sampling***. Further discussion of the decision to focus on exposure from the primary residence at the time of sampling is provided in Section G4.3 below in the sections on spatial and temporal variability.

G4.2 WHAT ADJUSTMENT FOR MEASUREMENT ERROR IS APPROPRIATE

The first question to be asked when addressing measurement error is: Is an adjustment for measurement error necessary? The appropriateness of an adjustment for measurement error is dependent on the use of the statistical model.

G4.2.1 Errors-In-Variables Adjustment To Model "True" Lead Exposure

A primary differentiation in model use concerns whether the model is being used to characterize the relationship between observed blood-lead levels in children and "true" lead exposures or whether the model is being used to predict blood-lead concentrations based on measured levels of environmental lead. The former case is the classic measurement error problem (Carroll, et al., 1995). Although this case may be of interest to EPA in documenting the extent of the lead problem, the primary use of the empirical model in the Section 403 rulemaking is for the latter case, prediction.

Therefore, because the empirical model is not intended to be used as a dose-response model, but rather is intended to be used to predict blood-lead levels based on measured levels of environmental lead, a classic errors-in-variables approach that would model the relationship between "true" lead exposure and children's blood lead concentrations was considered inappropriate for this analysis.

G4.2.2 Adjustment To Account For Differences In Measurement Error Between Dust Sampling Methods Used In The Rochester Study And Those Used In The HUD National Survey

In order to predict the national distribution of childhood blood-lead concentrations (prior to, and following implementation of Section 403 rules), the empirical model based on the Rochester Study must be combined with environmental data observed in a nationally representative sample (the HUD National Survey). As mentioned earlier, the dust and soil sampling methods were different between these two studies and therefore an adjustment for both systematic differences and differences in measurement error between the Rochester dust-lead and soil-lead predictor variables and the HUD National Survey dust-lead and soil-lead predictor variables must be considered.

An empirical model unadjusted for the effects of differences in the lead exposure predictor variables would be appropriate for prediction of the national distribution of blood-lead concentrations (prior to, and following Section 403 interventions) if the following four assumptions are met:

1. The sampling scheme for environmental lead implemented in the Rochester Lead-in-Dust Study (or other studies used for model building) is similar to the sampling scheme implemented in the HUD National Survey.

2. The sampling collection devices and instruments used to measure lead have similar properties with respect to measurement error between the Rochester Study and the HUD National Survey.

3. The distribution of observed environmental lead levels is similar between the Rochester Lead-in-Dust Study and the HUD National Survey.

4. The characteristics of the relationship between blood-lead and environmental lead in Rochester is the same as in the U.S. as a whole.

If either of the first two assumptions are not met, it would be necessary to adjust the model for differences in measurement error between variables constructed using the Rochester data and variables constructed using the HUD National Survey data. Although this can be considered an adjustment for "measurement error," the resulting model would not be interpreted as the relationship between blood-lead and "true" environmental lead levels (measured without error). Rather, this adjustment will account for differences in variability related to the different sampling methods to facilitate a more accurate prediction of the national distribution of childhood blood-lead concentrations.

If the third or fourth assumptions are not met, it raises the question as to whether the data from the Rochester Lead-in-Dust Study is an appropriate source of data for informing decisions concerning lead exposures nationwide.

Initial investigation of the data suggested that the first two assumptions were not met by the observed data in the two studies; and therefore, ***an adjustment for the differences between dust-lead and soil-lead predictor variables used in the model building process and dust-lead and soil-lead input variables from the HUD National Survey used in the prediction process is warranted.***

A related issue concerns the degree to which equation error (or an incorrect mathematical form of the model) can affect the accuracy and precision of model predictions. Measurement error and the form of the model are directly related in that the specific methodology for a measurement error adjustment is dependent on the form of the model.

G4.3 DEFINITION AND CHARACTERIZATION OF MEASUREMENT ERROR ASSOCIATED WITH EACH PREDICTOR VARIABLE

While it was determined that a classic adjustment for measurement error (Carroll, et al., 1995) was not appropriate for this particular use of the model, the statistical adjustment to the model for differences in sampling methods requires estimates of the variability associated with measuring the dust-lead and soil-lead exposure predictor variables. The following equation represents the three sources of variability that contribute to an estimate of measurement error in a dust-lead (or soil-lead) sample from the primary residence at the time of sampling and that are taken into account in the statistical adjustment to the model for differences in sampling methods:

$$\sigma^2_{\text{Measurement Error}} = \sigma^2_{\text{Spatial}} + \sigma^2_{\text{Sampling}} + \sigma^2_{\text{Laboratory}}$$

Another potential component of variability was temporal variability, $\sigma^2_{\text{Temporal}}$, but this component of variability was not included in any measurement error adjustments for reasons that are discussed below. The question of whether or not it is appropriate to consider any particular component of variation as part of the estimated measurement error for an exposure variable is dependant on the interpretation of the exposure variable and the way it is being used in the statistical model. Each component, including temporal variability, is discussed in the following subsections with respect to characterizing measurement error in the lead exposure predictor variables.

Details concerning the estimation of variability associated with measurement error in the dust-lead predictor variables are provided in Section G10.

G4.3.1 Spatial Variability

Spatial variability ($\sigma^2_{\text{Spatial}}$) represents variability in environmental lead levels among all possible locations on the surface(s) being tested as part of the sampling scheme. Although an ideal lead exposure variable would characterize lead-levels from all the surfaces which are related to a child's lead exposure (both inside and outside of the primary residence), the environmental data corresponding to a subject's lead exposure is usually limited to the sampling schemes implemented during a study. (For residential risk assessments, it is limited to the sampling schemes specified by the Section 402 rule.) It is an assumption that the sampling schemes that were implemented in these studies provide a sample of environmental lead as would be obtained in a risk assessment.

Lead measures outside the primary residence are unlikely to be taken in a risk assessment. There appear to be two ways of viewing lead exposures that occur outside the primary residence (such as in a day care center):

1. Lead exposure that occurs outside the primary residence is not captured by the observed lead exposure variables. Outside exposure represents a group of covariates that are not included in the statistical models, and therefore, $\sigma^2_{\text{Spatial}}$ would be limited to the variability of environmental lead that occurs among all possible locations within the primary residence.

2. Lead exposure that occurs outside the primary residence is captured by the observed lead exposure variables (measured within the primary residence), based on an assumption that levels of environmental lead inside the primary residence are similar to levels of lead found outside the primary residence. Under this assumption, the definition of $\sigma^2_{\text{Spatial}}$ would be expanded to include the variability of environmental lead that occurs among all possible locations to which a child has been exposed (both inside and outside the primary residence).

We accepted the first viewpoint of spatial variability ($\sigma^2_{Spatial}$) based on the following three facts:

1. There is no known information that can be used to verify the assumption that lead-levels in paint, dust, or soil within the primary residence are representative of lead-levels that occur outside the home.

2. There is no known information that can be used to estimate $\sigma^2_{Spatial}$ under an expanded definition which includes all surfaces to which a child is exposed (both inside and outside of the primary residence). However, there is information that can be used to estimate spatial variability in environmental lead levels that occur within a primary residence.

3. Environmental interventions that will occur under Section 403 will likely be focussed on reducing residential exposure to lead. It may therefore be inappropriate to develop a model in which the predictor variables are interpreted in a way which represents exposure that occurs outside of the primary residence.

Spatial variability was taken into account in the statistical adjustments to the model for differences in dust and soil sampling methods.

G4.3.2 Sampling Variability

Sampling variability $\sigma^2_{Sampling}$ represents variability introduced during the physical collection of environmental samples, and is a typical source of measurement error associated with the lead exposure predictor variables. Examples of variability that may be classified as sampling variability when collecting dust samples include:

- variability associated with sampling methods, e.g. wipe versus vacuum sampling
- variability associated with sampled surfaces, e.g. carpeted versus uncarpeted floors
- variability associated with properties of the given sample, e.g. particle size and dust-loading.

Examples of variability that may be classified as sampling variability when collecting soil samples include:

- variability associated with sampling methods, e.g. coring tool versus grab sample
- variability associated with sampled surfaces, e.g. bare soil versus covered soil
- variability associated with properties of the given sample, e.g. fraction of soil sample that is fine (versus coarse).

Sampling variability was taken into account in the statistical adjustments to the model for differences in dust and soil sampling methods.

G4.3.3 Laboratory Variability

Laboratory variability ($\sigma^2_{Laboratory}$) represents variability in the laboratory analysis of an environmental sample, and includes error in sample preparation and analytical error. It is often the case that laboratory error is a very small component of the total measurement error associated with a sample result.

Laboratory variability was taken into account in the statistical adjustments to the model for differences in laboratory methods for measuring lead in dust and soil samples.

G4.3.4 Temporal Variability

Temporal variability ($\sigma^2_{Temporal}$) represents the variability over time in environmental lead levels on the locations(s) selected to be part of the sample. Although lead levels in paint may not be subject to substantial temporal variability, it is documented that lead levels in dust and soil vary over time.

Since we are interpreting the lead exposure variables as being representative of current lead exposure (as would be measured in a Section 402 Risk Assessment) rather than cumulative lead exposure, temporal variability in environmental lead levels was not taken into account in the statistical adjustments to the model for differences in dust and soil sampling methods.

G5.0 MODEL BUILDING BASED ON DATA FROM THE ROCHESTER STUDY

This chapter describes the steps involved in the development of a multi-media predictive model based on data observed in the Rochester Lead-in-Dust Study. First, single media models of the Rochester data were investigated, then the variables identified from them were used to explore joint media models. Diagnostic analyses are described which were used to validate assumptions made during model development. Finally, information from these efforts was used to develop a multi-media predictive model based on data observed in the Rochester Study.

G5.1 USE OF SINGLE MEDIA MODELS (Bivariate Relationships Between Blood-Lead and Each Potential Variable)

Statistical modeling of the data from the Rochester Lead-in-Dust Study began with an initial evaluation of the bivariate relationship between blood-lead concentration and each individual exposure variable or select covariate. This evaluation included an assessment of all five candidate statistical models discussed in Section G3.

Section G11 contains for each potential exposure variable constructed from the Rochester Lead-in-Dust Study Data, a figure which displays the estimated regression curve for each candidate statistical model plotted along with the observed data, as well as a table which summarizes parameter estimates and associated standard errors for each candidate model. Note that parameter estimates and associated standard errors for the active/passive uptake model are not included in the tables in Section G11, because in most cases, the $F_{Passive}$ parameter was estimated as zero in the bivariate models, and thus, the active/passive uptake model reduces in form to the active uptake model. Candidate models and the strength of the relationship between blood-lead and each variable were compared using measures of R^2 and estimated likelihood ratios. R^2 (also called the coefficient of determination) is a measure of the proportion of the variability in childhood blood-lead concentrations that is explained by a model. Estimated likelihood ratios were calculated using parameter estimates from each model and the observed data. Use of the likelihood ratio as a diagnostic tool is discussed in Section G5.3 on regression diagnostics.

Results of the bivariate statistical analysis of the relationship between blood-lead concentration and each potential exposure variable from the Rochester Lead-in-Dust Study Data demonstrated the following:

1. The variables representing the presence and severity of interior deteriorating lead-based paint were significant predictors of blood-lead. The variables representing the presence and severity of exterior deteriorating lead-based paint were only borderline significant at the 0.05 level.

2. Measures of floor dust-lead loading from uncarpeted surfaces were better predictors of blood-lead than measures of floor dust-lead loading from carpeted surfaces.

3. Measures of dust-lead loading from window wells were better predictors of blood-lead than measures of dust-lead loading from window sills.

4. Both measures of soil-lead concentration (Dripline & Play-Area) were strong predictors of children's blood-lead concentration. Using Dripline soil Pb concentration (n=186) allowed more children/houses to enter the model versus Play-area (n=87).

5. Pica for paint chips was a significant predictor of blood-lead. Pica for soil was borderline significant.

6. The indicator variable representing race (black) was the strongest single predictor of blood-lead concentrations.

7. Age was not significantly associated with blood-lead in the Rochester data.

G5.2 DESCRIPTION OF JOINT MEDIA MODELS (Development of a Multimedia Exposure Statistical Model)

After assessing the bivariate relationships with each variable under consideration, the variables were systematically evaluated in an effort to develop a parsimonious multimedia exposure model for each source of data. There were a number of technical issues involved in the fitting of these models, including variable selection, collinearity among environmental exposure variables, and details concerning the use of non-linear regression:

G5.2.1 Variable Selection and Collinearity

Variable selection for the multimedia exposure model was based on several properties, including strength of relationship with blood-lead concentration as estimated using the bivariate statistical models, predictive power of each variable when included into a model with competing sources of lead exposure, and interpretability of the parameter estimates. Another goal related to variable selection was to develop a predictive model that was based on lead exposure from the three environmental media; paint, dust and soil. Thus, measures of lead exposure from paint, dust, and soil were considered as primary variables in the statistical analyses, and all other variables were considered as secondary variables. If a secondary variable was competing with a primary exposure variable in the multimedia exposure model (in terms of explaining variability in childhood blood-lead concentration), the secondary variable was excluded from the model in its final form.

Another issue in variable selection is the fact that the multimedia exposure models included variables which represent lead-levels in paint, dust, and soil from each residential unit. These measures tend to be correlated, and may result in meaningless parameter estimates when jointly added to the same statistical model (i.e. the association between blood-lead and environmental-lead might be estimated as negative for one or more sources of exposure in the

joint model). To avoid negative parameter estimates for lead exposure predictor variables, all five candidate models were originally fitted using non-linear regression models with constraints on the parameter estimates associated with exposure variables (the parameter estimates for these variables were constrained to be greater than or equal to zero). Log-linear models with positive parameter estimates for lead exposure predictor variables were later fitted using standard linear regression models. The models occasionally converged to local maximums rather than the global maximum likelihood solution, however, this problem was resolved by identifying improved starting values for each model. Further discussion of collinearity diagnostics is presented in Section G5.3 and Section G12.

G5.2.2 Multimedia Exposure Model Development

As discussed above, many combinations of variables were considered for the multi-media exposure model. Section G13 presents details of statistical model fittings for four sets of variables which met the variable selection criteria discussed above. The variable selection and model development work resulted in the following general conclusions:

1. Measures of soil-lead concentration from the dripline, dust-lead loading from floors, dust-lead loading from window sills, interior deteriorated lead based paint, pica for paint, and race were consistent predictors of blood-lead concentrations. Window sill lead loading appeared to compete with interior deteriorated lead-based paint as a predictor of blood-lead concentration.

2. A reduced set of variables (including measures of lead in paint, dust and soil, race and pica for paint) resulted in statistical models which were able to explain roughly 40% of the variability in children's blood-lead concentrations.

3. The log-additive model was outperformed by the other candidate models, as indicated by log likelihood statistics presented in Section G13, largely due to a saturation of the response at higher levels of environmental lead.

4. The $F_{Passive}$ parameter in the Passive/Active Uptake model was consistently estimated at or very close to zero. The Active Uptake model may therefore be a more appropriate model (since it won't be over-parameterized).

5. The log-linear model consistently outperformed all other candidate models (with the same variables) based on an evaluation of log likelihoods, as can be seen in Section G13.

Parameter estimates and associated standard errors of a series of four different multi-media exposure models (each of which included a different set of predictor variables) are provided in Section G13. Each table in Section G13 contains the results of fitting all five candidate statistical model forms to data from the Rochester Lead-in-Dust Study.

G5.3 REGRESSION DIAGNOSTICS

This section describes the diagnostic analyses performed as part of development of the multi-media predictive model using data from the Rochester Lead-In-Dust Study. Through the use of regression diagnostics the adequacy of fit of the various candidate models developed to the data observed can be determined, and model assumptions can be verified. For these models, the following regression diagnostic procedures were performed:

1. A normal quantile plot of the residuals was created. The normal quantile plot approximated a straight line indicating that residuals (errors) were approximately normally distributed, as assumed.

2. Residual values were plotted versus predicted values. This scatterplot did not indicate signs of nonconstant variance (if points spread out or tighten up as you move from left to right) or nonlinearity (if points look quadratic or bow-shaped). The scatterplot exhibited no pattern, indicating no such problems. Similarly, plots of residuals versus predictors indicated no discernible pattern.

3. Cook's distance and DFFITS (both measures of influence) were plotted versus studentized residuals (a measure of how far an observation deviates from the modeled relationship) to indicate potential outliers - points with undue influence and points lying far outside the model's prediction. These plots of Cook's distance and DFFITS were produced only for the log-linear models, which were implemented using standard linear regression, and identified no obvious outliers or influential points.

4. For a closer examination of how points influence model parameter estimates, the models were fit while excluding a single point at a time. Analysis of the coefficients adjusted for their standard error (intercept, and coefficients of PbS, PbF, PbW and PbP), including plots, again identified no major problems with influential data points.

5. Partial regression leverage plots were created for the environmental measures of lead exposure: dripline soil, floor dust from carpeted and uncarpeted floors, paint/pica hazard, and window sill dust. A partial regression leverage plot that exhibits a strong linear relationship between blood-lead and the variable under consideration is indicative of a strong linear relationship between blood lead and the environmental measure of lead exposure while controlling for all the other variables in the model. Partial regression leverage plots were produced only for the log-linear models, which were implemented using standard linear regression, and indicated an adjusted positive relationship for each lead exposure variable included in the multi-media predictive model.

6. Partial R^2 comparisons between predictor variables included in the model were calculated. A high partial R^2 indicates greater importance in predicting blood-lead concentration.

7. Estimated log-likelihoods were calculated using parameter estimates from each model and the observed Rochester data, and the likelihood ratios between different models were then assessed. The likelihood ratio (LR) is equivalent to the ratio of the data's probability under one model compared to its probability under a second model. The likelihood ratio evaluation consistently indicated that the log-linear model provided the best fit to the data.

8. An analysis into the effects of collinearity using several methods was conducted during the development of the multi-media predictive model. Estimates of the tolerance statistic and the variance inflation factor associated with each predictor variable in the model were calculated, along with a single value decomposition for the design matrix of observed predictor variables in the model. These analyses suggested that the model did not suffer from a problem with collinearity.

The above regression diagnostics and tests of collinearity among explanatory variables for the multi-media predictive model are provided in detail in Section G12. Based on the regression diagnostics on the multi-media predictive model it was concluded that:

- no influential or outlying points should be deleted from the analysis,
- the model developed fits the data observed,
- model assumptions are verified, and
- the model does not appear to suffer from a severe problem with collinearity.

G5.4 THE MULTI-MEDIA PREDICTIVE MODEL BASED ON ROCHESTER DATA

The criteria used for the selection of variables in the multi-media predictive model emphasized use of measures of environmental lead and other factors observed in both the Rochester Lead-in-Dust Study and the HUD National Survey. Variables whose definition provided a convenient translation when applied to the National Survey, whose predictive power in Rochester were high, and whose spread in the National Survey populations covered a wide enough range of values, were used in the empirical model. For example, the paint/pica variable was chosen for use in the multi-media predictive model because it was a better predictor and because application of the paint (75th percentile) variable in the HUD National Survey data resulted in a variable that provided very little discrimination between houses in the survey. Another example is that although the variable Bare_flr was a stronger predictor of blood-lead than the variable Floor_A in the Rochester Study, Floor_A was a more appropriate choice for construction in the HUD National Survey, and was therefore selected for use in the multi-media predictive model. Therefore, measures of lead in soil, floor dust, window sill dust and the paint/pica variable were chosen for use in the multi-media predictive model. The final mathematical form of this model was:

$$\ln(PbB) = \beta_0 + \beta_1 \cdot \ln(PbF) + \beta_2 \cdot \ln(PbW) + \beta_3 \cdot \ln(PbS) + \beta_4 \cdot PbP + e$$

where PbB represents the blood-lead concentration, PbF corresponds to measurements from interior floor dust, PbW represents environmental lead from window sills, PbS represents soil-lead, PbP represents paint hazard, and e represents the residual error left unexplained by the model. Parameter estimates and associated standard errors, and measures of R-squared and the residual standard deviation for the empirical model are provided in Table G-3. Note that the parameter estimate associated with floor dust-lead loading was only borderline statistically significant when considered jointly with the effect of window sill dust-lead loading (and other exposure variables) in the multi-media predictive model.

Table G-3. Parameter Estimates and (Associated Standard Errors) for the Multi-Media Predictive Model Based on Data from the Rochester Lead-in-Dust Study

Parameter	Variable Description	Estimate
β_0	Intercept	0.418 (0.240)
β_1	log (PbF): Area-Weighted Arithmetic Mean (Wipe) Dust-Lead Loading from Any Floor (Carpeted or Uncarpeted)	0.066 (0.040)
β_2	log (PbW): Area-weighted Arithmetic Mean (Wipe) Dust-Lead Loading from Window Sills	0.087 (0.036)
β_3	log (PbS): Dripline Soil-Lead Concentration (fine soil fraction)	0.114 (0.035)
β_4	PbP: Indicator of Interior Paint/Pica Hazard	0.248 (0.100)
R^2	Coefficient of Determination	21.67%
	Root Mean-Square Error (Residual Error)	0.56188

The above multi-media predictive model is used in the Section 403 Risk Assessment to determine the probability that a child in the Rochester Study exposed to specific levels of lead in paint, dust and soil will have a blood-lead concentration exceeding 10 µg/dL.

G6.0 THE EMPIRICAL MODEL

The goal of the empirical model is to provide a relationship between blood-lead concentration and various environmental lead exposures as measured in the HUD National Survey for use in the Section 403 risk assessment. Unfortunately, the HUD National Survey contains no information about blood-lead concentration. However, data from the Rochester Lead-in-Dust Study (i.e. the multi-media predictive model) can provide a basis for the empirical model. At issue is how to use the multi-media predictive model based on the Rochester data set to develop an empirical model applicable to the data observed in the HUD National Survey.

Matters are complicated by the fact that the sampling methodology used to measure lead exposures in HUD is different from that used in Rochester. Thus, some variables have a different interpretation in each of these two studies. Specifically, two of the lead exposure measurements in HUD are blue nozzle floor dust lead loading and blue nozzle window sill dust lead loading, compared to floor wipe dust lead loading and window sill wipe dust lead loading in Rochester. Another example is that the soil variable in Rochester was based on a composite sample from the dripline area adjacent to the house, whereas in the HUD National Survey, the soil variable was based on a weighted average of samples collected from dripline, entryway and remote locations (with weights of 25%, 25% and 50%, respectively). Also the paint/pica hazard predictor variable was constructed differently between the Rochester Study and the HUD National Survey data. The primary difference was that the paint/pica hazard input variable from the HUD National Survey data was based on the measures of paint on both interior and exterior surfaces, whereas the variable used in Rochester for estimation of the effect of paint/pica hazard was based on measure of paint on only interior surfaces. Lead based paint on deteriorated exterior surfaces was not considered in the estimation of the paint/pica model parameter based on Rochester data because approximately 84 percent of houses in the Rochester Study were built prior to 1940 and as a result virtually every home surveyed in the Rochester Study had lead based paint on exterior surfaces. Therefore, a paint/pica hazard variable which included presence of exterior lead based paint in Rochester lost its statistical significance and its predictive power. The differences in paint/pica variable construction between the Rochester and HUD National Survey is considered minor in comparison to the differences in dust and soil sampling methodologies. Table G-4 provides details comparing the construction and interpretation of variables in both the Rochester Lead-in-Dust Study and the HUD National Survey.

The following statistical method was used to account for differences in dust and soil sample collection methods between the Rochester Study and the HUD National Survey when assessing the impact of 403 rulemaking on children's blood-lead levels. The method involves establishing a relationship between blood-lead and environmental variables as measured by methods used in the Rochester Study (i.e. the multi-media predictive model based on Rochester Data), and then adjusting this relationship to use dust-lead and soil-lead variables as measured in the HUD National Survey. The adjustment takes into account both systematic differences and differences in error structures between the Rochester wipe dust-lead and drip-line soil-lead predictor variables versus the HUD National Survey Blue Nozzle dust-lead and averaged soil-lead predictor variables. The method provides a relationship between blood-lead concentration,

Table G-4. Variable Construction in the Rochester Lead-in-Dust Study and the HUD National Survey

Predictor Variable	Rochester Study	HUD National Survey Input Variables
Soil	Natural log transformation of dripline soil-lead concentration (fine soil fraction).	The natural log transformation of the weighted average of dripline, entryway and remote soil-lead concentrations, with weights of 25%, 25% and 50% respectively when all three soil samples were collected. If these values were missing, an imputed value[1] was used.
Floor Dust	The natural logarithm of the area weighted arithmetic average (wipe) dust-lead loading from carpeted and uncarpeted floors.	The natural logarithm of the area-weighted arithmetic average dust-lead loading (Blue Nozzle Vacuum) from 3 sample locations (wet, dry and entry rooms) was used as the measure of lead in dust. If the dust-lead loadings from all of the 3 sample locations were missing, an imputed value[1] was used.
Window Sill Dust	The natural logarithm of the area-weighted arithmetic average (wipe) dust-lead loading from window sills.	The natural logarithm of the area-weighted arithmetic average dust-lead loadings (Blue Nozzle Vacuum) from window sills from 2 sample locations (wet and dry rooms). If the window sill dust-lead loadings from both sample locations were missing, an imputed value[1] was used.
Interior Pica/Paint	An indicator variable which was nonzero when the following conditions each existed in a residential unit: presence of deteriorated or damaged interior paint; presence of interior lead-based paint; and presence of a child with paint pica. The paint variable had values of: 0 No LBP (XRF reading < 1), or condition[a] is Good, or child does not exhibit pica; 1 LBP (XRF reading ≥ 1), condition is Fair or Poor, and child exhibits pica rarely; 2 LBP (XRF reading ≥ 1), condition is Fair or Poor, and child exhibits pica at least sometimes.	HUD National Survey homes were determined to have deteriorated LBP whenever there is any deterioration in interior or exterior lead-based paint, as measured by square footage (that is, square footage of deteriorated LBP surface > 0). That is, the LBP indicator was defined as 1 Whenever square footage of surface exhibiting deteriorated LBP (interior and exterior) > 0 0 Otherwise The pica factor was only considered for houses with deteriorated LBP. In these houses, it was assumed that 9% of U.S. children aged 1-2 years have pica for paint. For the children with pica for paint, the pica value was defined to be 1.5[b].

[1] Imputed values for dust and soil were based on a presence of LBP indicator variable and on a house age-specific indicator. The presence of LBP indicator was defined as:

0 Predicted maximum XRF < 1 for both interior and exterior samples

1 Predicted maximum XRF ≥ 1 for either interior or exterior samples.

The house age-specific indicator had categories: Pre-1940, 1940-1960, 1960-1979, Post-1979. The imputed values for dust and soil were constructed by taking the means for the associated subsets formed by crossing the paint and age of house categories.

[a] Condition of the paint in the Rochester Lead-in-Dust Study is described in Table G-2.

[b] The Paint/Pica Hazard Variable was described in Table G-2. A value of 1.5 was chosen as the input value for those children exhibiting pica in applying the empirical model to the HUD National Survey.

floor and window sill dust-lead loadings, soil-lead concentrations, and other covariates as observed in the HUD National Survey. An errors in variables measurement error adjustment is applied as an intermediate step in reaching this goal. The method for adjusting the multi-media predictive model may be described as follows, and is provided with complete detail in Appendix G1.

The first step involves fitting an errors in variables measurement error adjusted multi-media exposure model that assumes blood-lead concentration is a function of true unobserved floor and window sill dust-lead loadings and dripline soil-lead concentrations along with other covariates (paint/pica hazard) used in the model. While the dependence of blood-lead concentration on true dust-lead loadings, dripline soil-lead concentrations, and other covariates can not be observed, they can be estimated via equations (2.2.12) and (2.2.16) in Fuller, 1987. In order to use these equations for estimating this relationship, the measurement error associated with each particular dust-lead loading and soil-lead concentration must be obtained. This is achieved by taking individual measurements of dust-lead loadings and soil-lead concentrations within households and calculating their variability. The average of all within household variances is then used as an estimate of the true measurement error associated with each particular dust-lead loading and soil-lead concentration. The estimated measurement errors are then used to calculate parameter estimates for a model based on Rochester data that relates blood-lead concentration to true dust-lead loadings, dripline soil-lead concentrations, and other covariates (paint/pica hazard). Keep in mind that the model must be developed using Rochester data because there is no blood-lead concentration variable in the HUD data set.

If the goal had been to identify the nature of the dependence of blood-lead concentration on true floor and window sill dust-lead loadings, dripline soil-lead concentrations, and the other covariates, then the adjustment described above would have been all that was required. However, the relationship of interest is blood-lead concentration as a function of floor and window sill dust-lead loadings, average soil-lead concentrations, and other covariates (paint/pica hazard) as observed in the HUD National Survey. Therefore, adjusting for measurement error is only the first step toward a final solution to this problem.

The next step in this process is to define the relationship between blood-lead concentrations, observed dust-lead and soil-lead predictor variables as measured in both Rochester and HUD, dust-lead and soil-lead predictor variables measured without error on the scale of measure used in Rochester, and any other covariates (paint/pica hazard) in the multimedia exposure model. It is assumed that these random variables jointly follow a multivariate normal distribution. Standard statistical theory then allows for deriving the distribution of blood-lead concentration conditioned on floor and window sill dust lead loadings, average soil lead concentrations, and other covariates as measured in HUD. Estimates of the parameters for a multimedia exposure model that relates blood-lead concentration to lead exposures as measured in the HUD National Survey are obtained from this conditional distribution.

The final step in developing the empirical model was to derive an estimate for the intercept. The empirical model intercept was designed to calibrate the model so that the predicted national (pre-403) geometric mean blood-lead concentration obtained from applying the empirical

model to data observed in the HUD National Survey equals the geometric mean blood-lead concentration estimated in Phase 2 of NHANES III.

The empirical model involves an adjustment to the multi-media predictive model based on the Rochester Study to allow use of Blue-Nozzle dust-lead loadings rather than wipe dust-lead loadings and average soil-lead concentration rather than dripline soil-lead concentration. The final mathematical form of this model is:

$$\ln(PbB) = \beta_0 + \beta_1 \cdot \ln(PbF_{BN}) + \beta_2 \cdot \ln(PbW_{BN}) + \beta_3 \cdot \ln(PbS) + \beta_4 \cdot PbP + e$$

where PbB represents the blood-lead concentration, PbF_{BN} and PbW_{BN} correspond to dust-lead loading from interior floors and window sills respectively (for samples collected in the HUD National Survey with the blue nozzle vacuum), PbS represents average soil-lead concentration, PbP represents paint/pica hazard, and e represents the residual error left unexplained by the model. Table G-5 provides parameter estimates and associated standard errors for the empirical Model developed to predict the national distribution of children's blood-lead concentrations using data as observed in the HUD National Survey. The standard errors provided in Table G-5 were estimated using a Bootstrap Algorithm which is detailed in Section G10.4.

Table G-5. Parameter Estimates and Associated Standard Errors for the Empirical Model used to Predict the National Distribution of Children's Blood-Lead Concentration Based on Data from the HUD National Survey

Variable	Parameter	Estimate (Standard Error)
Intercept	β_0	0.650 (0.154)
Floor Dust-Lead Loading (Blue Nozzle Vacuum)	β_1	0.032 (0.044)
Window Sill Dust-Lead Loading (Blue Nozzle Vacuum)	β_2	0.050 (0.031)
Average Soil-Lead Concentration	β_3	0.094 (0.043)
Paint/Pica Hazard	β_4	0.256 (0.098)
Error	$^2_{Error}$	0.313

G7.0 ESTIMATING THE NATIONAL DISTRIBUTION OF BLOOD-LEAD USING THE EMPIRICAL MODEL

As stated previously, the empirical model will be used in the Risk Assessment to predict a national distribution of children's blood-lead concentrations both before and after interventions resulting from the Section 403 standards. A nationally representative sample of environmental conditions in housing is required as input to the empirical model to predict a national distribution of children's blood-lead concentrations. The HUD National Survey is a nationally representative study which assessed environmental lead-levels in paint, dust and soil in residential housing. Environmental conditions observed in the HUD National Survey were used as input to the EPI model for predicting blood-lead levels in children 1-2 years old. A population of children aged 1-2 years is both the target age group for EPA's Risk Assessment, and the age group that was recruited in the Rochester Lead-in-Dust Study (thus the empirical model is representative of children in this age group). The empirical model is used to estimate an average log-transformed childhood blood-lead concentration associated with each home in the HUD National Survey.

As noted in Table G-5, the variables used for prediction are average soil-lead concentration, blue-nozzle vacuum dust-lead loading on floors (carpeted or uncarpeted), blue-nozzle vacuum dust-lead loading on window sills, and an indicator of paint/pica hazard. These variables, constructed from observed levels of lead in each HUD National Survey residential unit, are used as input to the empirical model for predicting the pre-403 national distribution of children's blood-lead concentrations.

To predict a post-403 national distribution of children's blood-lead concentrations, the following method was used to prepare the HUD National Survey Data for input into the empirical model:

[1] Observed levels of lead in environmental variables in the HUD National Survey were compared to proposed section 403 standards. Blue-nozzle vacuum floor and window sill dust-lead loadings were converted to wipe dust-lead loadings before comparison to the 403 standards.

[2] Section 403 interventions were triggered in HUD National Survey residential units that had levels of lead in environmental variables that were above the proposed standard. If an intervention was triggered, assumed post-intervention lead levels in environmental variables were substituted for observed levels according to the Section 403 risk assessment assumptions. Post intervention dust-lead levels that were specified in terms of wipe dust-lead loadings were converted to a blue nozzle vacuum scale for use in the prediction.

The distribution of blood-lead concentrations associated with each home was characterized by assigning a geometric mean (predicted by the empirical model) and a geometric standard deviation. A geometric standard deviation of 1.6 was assumed for the distribution of blood-lead concentrations associated with each home. The default geometric standard deviation of blood-lead concentrations for children at similar environmental-lead levels for the IEUBK

model is 1.6 and the estimated variability from the multi-media predictive model based on the Rochester Data was 1.76 as measured by the exponentiation of the root mean square error. Thus, a population of children (aged 1-2 years) associated with environmental lead levels found at each home in the HUD National Survey was constructed using the geometric mean blood-lead concentration predicted by the empirical model, an assumed geometric standard deviation of 1.6, and population weights based on the 1993 American Housing Survey adjusted to 1997.

The predicted national distribution of blood-lead concentrations can be characterized using a geometric mean and a geometric standard deviation. The predicted national geometric mean is calculated by taking a weighted geometric mean of the empirical Model predicted blood-lead concentration associated with each home in the HUD National Survey, using the adjusted weights for 1997. The predicted national geometric standard deviation is calculated by taking the square root of the sum of the predicted between-house variability and the assumed within-house variability. The predicted between-house variability is estimated as a weighted geometric variance among the empirical Model predicted blood-lead concentration associated with each home in the HUD National Survey, using the adjusted weights for 1997. Thus, the between-house variability represents the variability among the predicted blood-lead concentrations associated with the environmental conditions observed in each home in the HUD National Survey. The assumed within-house variability was $(1.6)^2$, and represents the expected variability among children who are exposed to similar environmental conditions. The predicted national geometric standard deviation relies on an assumption that the between-homes distribution of blood-lead concentration is log-normally distributed.

The predicted national distribution of children's blood-lead concentrations can also be characterized using exceedance percentiles (i.e. the percentage of children estimated to have blood-lead concentrations above a specified level, such as 10, 20 and 30 µg/dL). These exceedance proportions were calculated in two ways, first by using normal probability theory combined with the estimated national geometric mean and standard deviation, and second by empirical evaluation of a national population built by summing discretized populations of children associated with each home.

The second approach is robust to deviations from the assumed log-normal distribution of blood-lead concentrations between homes, and can be described as follows:

A distribution of blood-lead concentrations is constructed for each home using the empirical Model predicted geometric mean and the assumed within house geometric standard deviation of 1.6. Each of these distributions are then partitioned into seven discrete blood-lead intervals. Table G-6 provides the specific method for partitioning a distribution of log blood-lead concentrations into the seven intervals about the log of the geometric mean (predicted from the empirical model). Figure G-2 graphically illustrates this partitioning. The two tails of the distribution represent log blood-lead concentrations below or above 2.5 standard deviations from the mean, respectively. The percentage of the distribution assigned to each of these intervals, 0.62%, is based on the area under a standard normal curve for z-values less than -2.5 in the lower tail or greater than 2.5 in the upper tail. The assigned log blood-lead concentration for the lower tail is the expected value of a standard normal random deviate lying in the interval from $-\infty$ to -

2.5; the assigned log blood-lead concentration was similarly chosen for the upper tail, and mid-points were used for the finite-length intervals. The assigned blood-lead concentration for each interval was obtained by exponentiating the assigned log blood-lead for the interval. For example, for the lower tail,

$$e^{\mu - 2.82\times\sigma} = e^{\mu} \times e^{-2.82\times\sigma} = GM \times GSD^{-2.82} = \frac{GM}{GSD^{2.82}}$$

Table G-6. Allocation of Blood-Lead Distribution to Seven Intervals

Log Blood-Lead Concentrations			Assigned Blood-Lead Concentration for Interval
Interval for Log Blood Lead[a]	Percentage of Distribution in Interval	Assigned Log Blood Lead for Interval	
[-∞, μ - 2.5 *]	0.0062	μ - 2.82 * [b]	GM/[GSD$^{2.82}$]
[μ - 2.5 * , μ - 1.5 *]	0.0606	μ - 2.00 *	GM/[GSD$^{2.00}$]
[μ - 1.5 * , μ - 0.5 *]	0.2417	μ - 1.00 *	GM/[GSD$^{1.00}$]
[μ - 0.5 * , μ + 0.5 *]	0.3830	μ	GM
[μ + 0.5 * , μ + 1.5 *]	0.2417	μ + 1.00 *	GM*[GSD$^{1.00}$]
[μ + 1.5 * , μ + 2.5 *]	0.0606	μ + 2.00 *	GM*[GSD$^{2.00}$]
[μ + 2.5 * , + ∞]	0.0062	μ + 2.82 * [c]	GM*[GSD$^{2.82}$]

[a] Blood-lead concentrations were assumed to have a log-normal distribution with the geometric mean (GM) predicted by the empirical model and a geometric standard deviation (GSD) of 1.6 (the default geometric standard deviation for the IEUBK model). The distribution of log blood-lead concentrations was assumed to be normal with mean μ given by log(GM) and standard deviation given by log(GSD = 1.6).

[b] The expected value of a normal random deviate known to lie in the interval [-∞, -2.5] is -2.82.

[c] The expected value of a normal random deviate known to lie in the interval [2.5, + ∞] is + 2.82.

For this lower tail, if N children were associated with the specific housing condition (according to weights in the 1993 American Housing Survey adjusted to 1997) then 0.62 percent of the N children were assigned a blood-lead concentration of GM/GSD$^{2.82}$. The remaining 99.28 percent were similarly assigned to the other blood-lead concentrations presented in Table G-5 using the percentages given in the second column of the table. In this manner, the distribution of blood-lead concentrations of the N children were allocated to a distribution of blood-lead concentrations centered around the GM predicted by the empirical model with a GSD of 1.6. The predicted distributions at each housing condition were then combined to generate a distribution of childhood blood-lead levels over all of the housing conditions present in the HUD National Survey.

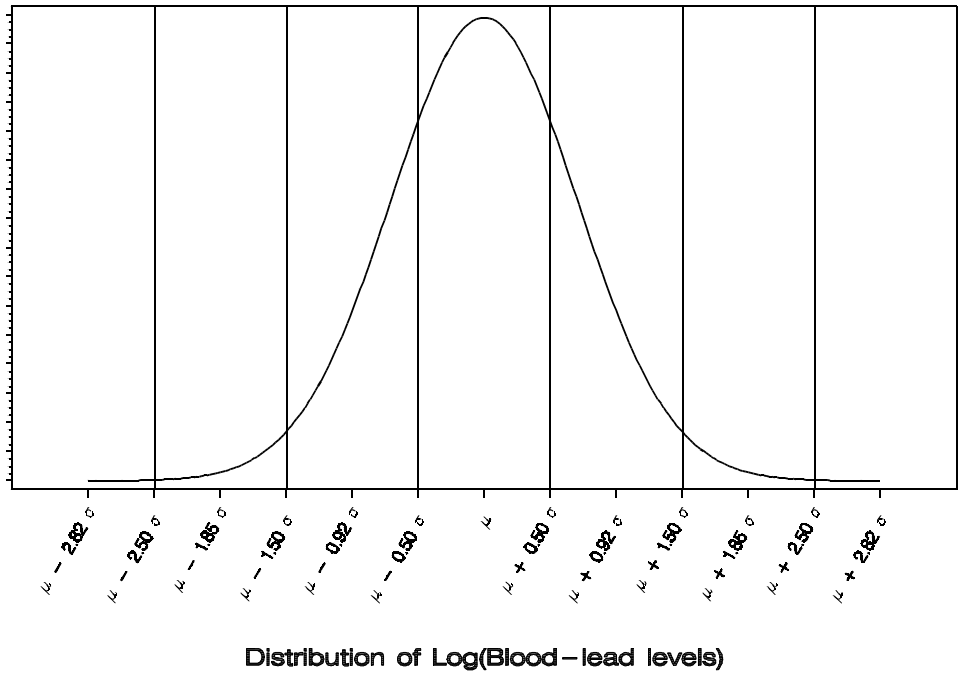


Figure G-2. Distribution of Blood-Lead Levels About Geometric Mean on Logarithmic Scale.

The exceedance percentiles can then be assessed by empirically tabulating the proportion of children in this constructed distribution who are above the target blood-lead concentrations of 10, 20 and 30 µg/dL.

G7.1 RESULTS OF THE COMPARISON WITH NHANES III

The predicted distribution of blood-lead concentrations obtained by applying the empirical model to the HUD National Survey Data was compared to NHANES III as a check on how well the empirical model performed. Table G-7 contains characteristics of the predicted blood-lead distribution for the empirical model, including estimates of exceedance proportions (the estimated proportion of blood-lead concentration exceeding 10, 20 or 30 µg/dL), the geometric mean, and the geometric standard deviation. Results in Table G-7 for the NHANES III distribution, the distribution of children recruited into the Rochester Lead-in-Dust Study, and the predicted national distribution based on applying the empirical model to data from the HUD National Survey (both before and after Section 403 interventions take place) are presented first with exceedance proportions calculated from the discretized distribution and second for exceedance proportions calculated assuming a log-normal distribution with the calculated geometric mean and geometric standard deviation.

Table G-7. Predicted National Distribution Characteristics for Empirical Model Compared to Rochester and NHANES III

Predicted Model Results	Parameter	Pre-Intervention Blood-lead Levels			Post-Intervention Blood-lead Levels[1]
		NHANES III	Rochester Study	Empirical Model	Empirical Model
National Geometric Mean	μ_P	3.14	6.36	3.14	3.03
National Geometric Standard Deviation	$_P$	2.09	1.85	1.71	1.67
Discretized Distribution Exceedance Percentiles (% of Population ≥ 10, 20 & 30 µg/dL)	% ≥10 µg/dL	5.88%	22.90%	0.00%	0.00%
	% ≥20 µg/dL	0.43%	2.90%	0.00%	0.00%
	% ≥30 µg/dL	0.07%	1.00%	0.00%	0.00%
Log-Normal Distribution Exceedance Percentiles (% of Population ≥ 10, 20 & 30 µg/dL)	% ≥10 µg/dL	5.75%	23.10%	1.54%	1.00%
	% ≥20 µg/dL	0.59%	3.13%	0.03%	0.01%
	% ≥30 µg/dL	0.11%	0.01%	0.0013%	0.0004%

[1] For illustration of a calculation of a post-intervention blood-lead distribution, standards were set at: 100 µg/ft^2 for floor dust-lead loading (wipe), 500 µg/ft^2 for window sill dust-lead loading (wipe), 2000 µg/g for soil removal, 5 ft^2 damaged LBP for paint repair, and 20 ft^2 damaged LBP for paint abatement. Post-403 lead levels for homes that were above the standard were adjusted to 40 µg/ft^2 for floor dust-lead loading (wipe), 100 µg/ft^2 for window sill dust-lead loading (wipe), 150 µg/g for soil removal, and 0 ft^2 damaged LBP for paint repair or abatement.

The results of the comparison with NHANES III for the revised empirical model indicate:

- The national geometric mean blood-lead concentration (pre-intervention) was calibrated to the geometric mean reported in NHANES III.
- The variability in the national distribution of blood-lead concentration predicted by the empirical model using the HUD National Survey (pre-403) is estimated at 1.71 (GSD), in contrast to a GSD of 2.09 for NHANES III.
- The estimated proportions of blood-lead concentrations of at least 10, 20, or 30 µg/dL using the empirical model predictions are much lower than the corresponding proportions estimated by NHANES III.

It should also be noted that NHANES III itself is only an estimate of the true national distribution of blood-lead concentrations (pre-403), and that an "exact" match of NHANES III does not mean an exact match of the true national distribution, nor does it guarantee that the model is appropriate for predicting a post-403 national distribution.

G8.0 DISCUSSION

The primary limitation associated with the Rochester Study is concern over the degree to which the Rochester Study may be considered representative of the nation as a whole. Differences between the Rochester Study population and the national population include the following:

a. Almost one-quarter (22.9%) of the Rochester children had observed blood-lead concentrations above 10 µg/dL, whereas only 5.9% of children aged 1-2 years nationwide were estimated to have blood-lead concentrations above 10 µg/dL by Phase 2 of NHANES III.

b. The geometric mean blood-lead concentration in Rochester is 6.4, whereas the geometric mean blood-lead concentration nationwide as estimated by NHANES III is 3.1. The GSD for Rochester is 1.9, compared to 2.1 for NHANES III.

c. Approximately 84 percent of the housing included in the Rochester Study was built prior to 1940, and there is a well documented relationship between age of housing and presence of lead-based paint. Only approximately 20% of housing nationwide was built prior to 1940.

d. Approximately 40% of the sample of children in the Rochester Study were African Americans, compared to an estimated 13% of the population of children nationwide (from 1997 US Census Projections), and compared to approximately 7% in the HUD National Survey.

e. Environmental levels of lead in soil in the Rochester Study were higher than would be expected in the HUD National Survey. For example, the geometric mean dripline soil-lead concentration in the HUD National Survey was approximately 75 ppm whereas the Rochester geometric mean was approximately 730 ppm.

f. Subjects recruited into the Rochester Study represent children whose primary exposure to lead was from dust, soil and paint at the primary residence. Children whose parents had lead exposure, who spent time away from the home, or whose homes underwent renovation or remodeling were excluded from the study. Only 376 of 1,536 families were eligible to participate in the study after the initial telephone screening. The selection criteria utilized in the Rochester Study may have resulted in a biased sample of children, since children who had potential lead exposure outside of the primary residence were excluded.

The difference in the observed blood-lead distributions between the Rochester Study and NHANES III is illustrated in Figure G-3. Although there are limitations associated with the Rochester Study, there are also positive aspects of the study that recommend its use:

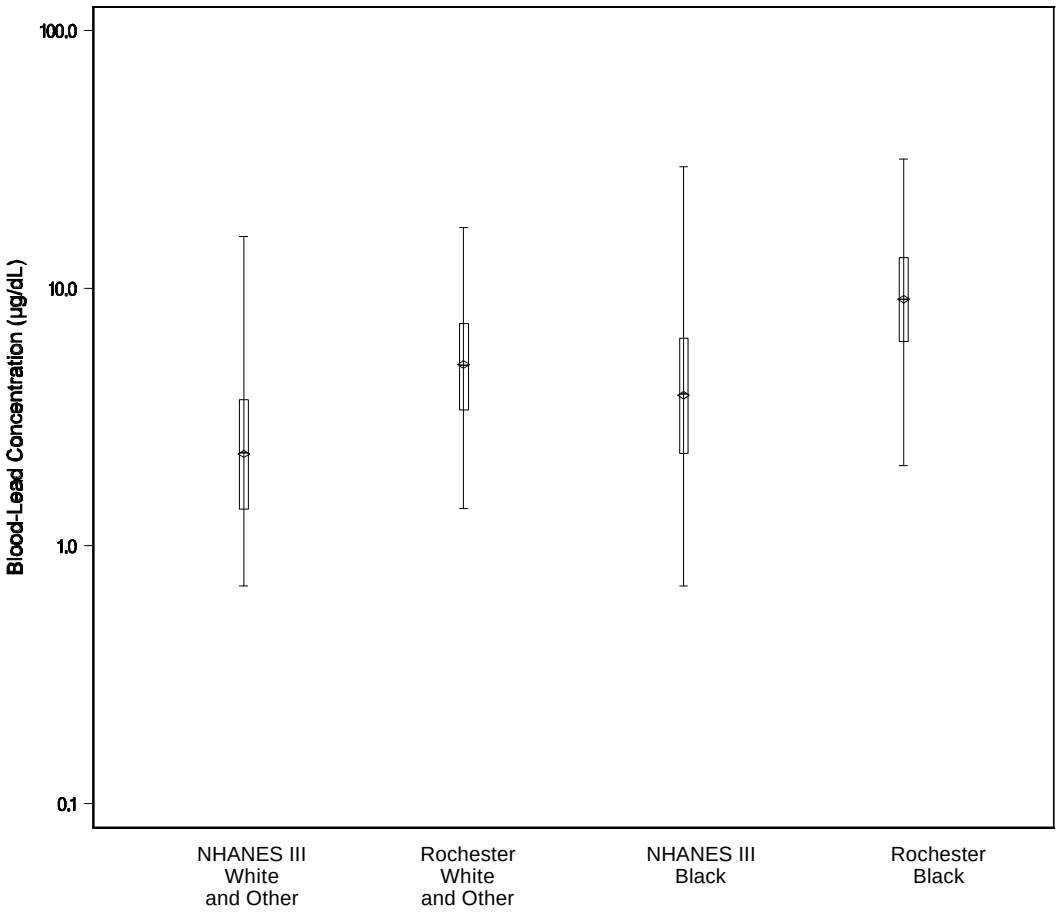


Figure G-3. Box Plot of Blood-lead Concentrations for Children Aged 1-2 Years for Phase II of NHANES III versus Rochester Data Sets.

a. all media, locations, and surfaces that are being considered for Section 403 standards were measured for lead in the Rochester Study.

b. the Rochester Study includes dust-lead loadings from wipe sampling and the Section 403 dust standard is expected to be based on dust-lead loading from wipe sampling.

c. the selection of homes and children in the Rochester Study, although targeted, was more random and more representative of a general population than is the case with most recent epidemiological studies of lead exposure in non-smelter communities.

The ability of an empirical model to predict the national distribution of blood-lead concentrations following Section 403 lead hazard reduction activities may be most severely limited by factors that are not included in the model. Reflecting its use in the Section 403 Risk Assessment, the empirical model accounts only for factors related to environmental lead

exposures at the residence, and does not account for other factors that might affect childhood blood lead. Such factors that may affect children's blood-lead concentration but may not be able to be controlled by the Section 403 rule include:

(1) home and personal cleaning habits,
(2) diet and nutritional status,
(3) bio-availability of the lead found in residential environmental media,
(4) non-residential exposures,
(5) inhalation exposure,
(6) children's behavior,
(7) socio-economic factors,
(8) renovation and remodeling (R&R) activity,
(9) hobbies,
(10) occupation.

Finally, it should be noted that the empirical model contains variables that differ from variables created for a best-fit of the Rochester data, because the goal of the empirical model was to provide a basis for using measures of lead from the HUD National Survey to predict a national distribution of childhood blood-lead concentrations. In particular, the empirical model differs from the multimedia regression model used to characterize the dose-response relationship between environmental-lead and blood-lead.

G9.0 REFERENCES

See Chapter 7 of Volume I for references cited within this appendix.

G10: Appendix on Methodology for Adjusting for Different Sampling Methods

Statistical Details

Section G10 of Appendix G is comprised of four sections that describe the statistical details associated with the Empirical Model. Section G10.1 explains statistical methodology used to account for differences in sample collection methods used in the Rochester Lead-in-Dust Study and the HUD National Survey. Section G10.2. describes the classic errors in variables regression model. Section G10.3 provides details on the estimation of variance components used as input to the above two statistical models. Finally, Section G10.4 explains the bootstrap algorithm used for approximating the standard errors associated with parameter estimates of the model that accounts for differences in sampling methods.

G10.1 STATISTICAL ADJUSTMENTS TO ACCOUNT FOR DIFFERENCES IN SAMPLE COLLECTION METHODS USED IN THE ROCHESTER LEAD-IN-DUST STUDY AND THE HUD NATIONAL SURVEY

The goal of this section is to provide a statistical methodology for adjusting the multi-media predictive model to appropriately use environmental lead levels observed in the HUD National Survey as inputs to the model. The adjustment takes into account both systematic differences and differences in error structures between the Rochester predictor variables and the HUD National Survey predictor variables. The method provides a relationship between blood-lead concentration and a set of lead exposure variables and other covariates as they were measured in the HUD National Survey. As an initial overview the method may be described as follows. Assume:

- Y represents children's blood-lead levels,
- R represents wipe dust lead loading observed in the Rochester Study,
- H represents blue nozzle dust lead loading observed in the HUD National Survey, and
- C represents covariates of interest which appear both in the Rochester Study and in the HUD National Survey.

The density of interest is children's blood lead levels as a function of lead exposures measured in the HUD National Survey, namely

$$F_{Y|H,C}(y|h,C)= \int F_{Y|R,H,C}(y|r,h,C) \cdot F_{R|H,C}(r|h,C) \ .$$

Given that we do not have a source of data with Y,R,H and C observed simultaneously, the method used for estimating $F_{Y|H,C}(y|h,C)$ is:

$$F_{Y|H,C}(y|h,C) = \int F_{Y|X,C}(y|x,C) \cdot F_{X|H,C}(x|h,C)$$

where X is a latent variable that represents dust lead loading measured without error.

This method assumes that Y can be modeled as a function of X using an errors-in-variables approach.

Details of the method are provided in the following subsections. Section G10.1.1 presents the methodology for the specific case of an errors-in-variables adjustment of a single covariate. This section is provided to aid in the understanding of the theoretical development of the model parameters. Section G10.1.2 presents the methodology for the general case of an errors-in-variables adjustment of one or more covariates. The Empirical model involves an errors-in-variables adjustment of three covariates: floor wipe dust lead loading, window sill wipe dust lead loading, and drip-line soil lead concentration. Thus, the Empirical model parameter development follows the methodology detailed in Section G10.1.2.

G10.1.1 MODELING BLOOD-LEAD AS A FUNCTION OF ONE VARIABLE MEASURED WITH ERROR AND OTHER SELECT COVARIATES.

The following theoretical development of the Empirical model parameters is specific to an errors-in-variables adjustment of a single covariate. Details are given for this specific case for two reasons:

1. The original theory was developed in this context.
2. The theoretical development is easiest to follow for a single variable adjustment.

In general, the theory applies to errors-in-variables adjustments for any number of covariates in the model. Section G10.1.2 below uses matrix notation to present the general theoretical details, which includes as a special case the errors-in-variables adjustment of a single covariate.

Definitions and Assumptions

Define the following variables:

- Y = The response variable, log of blood-lead concentration.
- R = Log of the area weighted arithmetic mean of the floor wipe dust lead loading as observed and measured in the Rochester Lead-in-Dust Study.
- H = Log of the area weighted arithmetic mean of the blue nozzle floor dust lead loading as observed and measured in the HUD National Survey.
- X = Log of the "true" unobserved area weighted arithmetic mean floor wipe dust lead loading (measured without error).
- C = A vector (or scalar) of remaining covariates used as independent variables. For the model detailed in this section, which adjusts for the measurement error in floor wipe dust lead loading only, C is a vector consisting of the variables drip-line soil lead concentration and paint/pica hazard. These covariates are assumed to be measured using identical methods in the Rochester Lead-in-Dust Study and the HUD National Survey.

The model assumes

(A)

$$\begin{bmatrix} Y \\ R \\ H \\ X \\ C \end{bmatrix} \sim N \left\{ \begin{bmatrix} \mu_Y \\ \mu_R \\ \mu_H \\ \mu_X \\ \mu_C \end{bmatrix}, \begin{bmatrix} \sigma^2_Y & \sigma^2_{YR} & \sigma^2_{YH} & \sigma^2_{YX} & \sigma^2_{YC} \\ & \sigma^2_R & \sigma^2_{RH} & \sigma^2_{RX} & \sigma^2_{RC} \\ & & \sigma^2_H & \sigma^2_{HX} & \sigma^2_{HC} \\ & & & \sigma^2_X & \sigma^2_{XC} \\ & & & & \sigma^2_C \end{bmatrix} \right\} ,$$

(B)

$$\begin{aligned} Y &= \alpha_{Y|X,C} + \beta_{Y|X(C)} Xf + \beta_{Y|C(X)} C + e_{Y|X,C} & e_{Y|X,C} &\sim N(0,\ \sigma^2_{Y|X,C}) \\ R &= X + e_{R|X} & e_{R|X} &\sim N(0,\sigma^2_{R|X}) \\ H &= \alpha_{H|X} + \beta_{H|X} X + e_{H|X} & e_{H|X} &\sim N(0,\sigma^2_{H|X}) \\ X &= \alpha_{X|C} + \beta_{X|C} C + e_{X|C} & e_{X|C} &\sim N(0,\sigma^2_{X|C}) \\ C &= \mu_C + e_C & e_C &\sim N(0,\sigma^2_C) \quad , \end{aligned}$$

and all errors are independent of one another.

The parameters $\alpha_{Y|X,C}$, $\beta_{Y|X(C)}$, and $\beta_{Y|C(X)}$ represent the intercept and slopes, respectively, associated with a regression of Y on X(unobserved) and C; and $\sigma^2_{Y|X,C}$ is the variability in Y unexplained by X and C. $\sigma^2_{R|X}$ is the measurement error associated with wipe floor dust lead loading in the Rochester Study. $\sigma^2_{H|X}$ is the measurement error associated with blue nozzle vacuum floor dust lead loading in the HUD National Survey. $\alpha_{H|X}$ represents a location shift in the distribution of H relative to the distribution of X. Similarly, $\beta_{H|X}$ represents a scale shift in the distribution of H relative to the distribution of X. $\alpha_{X|C}$ and $\beta_{X|C}$ are the intercepts and slopes, respectively, associated with a regression of X on the covariates in C. $\sigma^2_{X|C}$ represents the variability in $\underline{X}$ unexplained by the covariates in C.

In addition, the calculations that follow rely heavily on the assumption that the conditional distribution of X given C ($X|C$) is the same in the Rochester Lead-in-Dust Study and the HUD National Survey. This assumption will be referred to as an assumption of transportability.

Parameter Development

Using assumption (A) of Section G10.1.1, normal distribution theory implies that Y conditioned on H and C is normally distributed with the following parameters:

$$\mu_{Y|H,C} = \mu_Y + [\sigma^2_{YH} \quad \sigma^2_{YC}] \begin{bmatrix} \sigma^2_H & \sigma^2_{HC} \\ & \sigma^2_C \end{bmatrix}^{-1} \begin{bmatrix} H - \mu_H \\ C - \mu_C \end{bmatrix}$$

$$\sigma^2_{Y|H,C} = \sigma^2_Y - [\sigma^2_{YH} \quad \sigma^2_{YC}] \begin{bmatrix} \sigma^2_H & \sigma^2_{HC} \\ & \sigma^2_C \end{bmatrix}^{-1} \begin{bmatrix} \sigma^2_{YH} \\ \sigma^2_{YC} \end{bmatrix}$$

Solving the inverse matrix above and using assumption (B) of Section G10.1.1 for substitutions yields:

$$\mu_{Y|H,C} = \mu_Y + \begin{bmatrix} \dfrac{\beta_{H|X}\ \beta_{Y|X(C)}\ \sigma^2_{X|C}}{\beta^2_{H|X}\ \sigma^2_{X|C} + \sigma^2_{H|X}} \\ \dfrac{\beta_{X|C}\ \beta_{Y|X(C)}\ \sigma^2_{H|X}}{\beta^2_{H|X}\ \sigma^2_{X|C} + \sigma^2_{H|X}} + \beta_{Y|C(X)} \end{bmatrix}^T \begin{bmatrix} H - \mu_H \\ C - \mu_C \end{bmatrix}$$

$$\sigma^2_{Y|H,C} = \sigma^2_{Y|X,C} + \frac{\beta^2_{Y|X(C)}\ \sigma^2_{H|X}\ \sigma^2_{X|C}}{\beta^2_{H|X}\ \sigma^2_{X|C} + \sigma^2_{H|X}}$$

Using (B), observe that

$$(C) \qquad \beta^2_{H|X}\ \sigma^2_{X|C} + \sigma^2_{H|X} = \sigma^2_{H|C} \quad ,$$

where the left-hand side of (C) represents the portion of σ^2_H that remains after conditioning on C.

From (C),

$$\beta_{Y|H(C)} = \frac{\beta_{H|X}\ \beta_{Y|X(C)}\ \sigma^2_{X|C}}{\beta^2_{H|X}\ \sigma^2_{X|C} + \sigma^2_{H|X}} = \frac{\beta_{Y|X(C)}}{\beta_{H|X}} \left(\frac{\beta^2_{H|X}\ \sigma^2_{X|C}}{\beta^2_{H|X}\ \sigma^2_{X|C} + \sigma^2_{H|X}} \right) = \frac{\beta_{Y|X(C)}}{\beta_{H|X}} \left(1 - \frac{\sigma^2_{H|X}}{\sigma^2_{H|C}} \right) \quad ,$$

$$\beta_{Y|C(H)} = \beta_{Y|C(X)} + \frac{\beta_{X|C}\ \beta_{Y|X(C)}\ \sigma^2_{H|X}}{\beta^2_{H|X}\ \sigma^2_{X|C} + \sigma^2_{H|X}}$$

$$= \beta_{Y|C(X)} + \left\{ \beta_{X|C}\ \beta_{Y|X(C)} \left(\frac{\sigma^2_{H|X}}{\sigma^2_{H|C}} \right) \right\} \quad ,$$

and

$$\sigma^2_{Y|H,C} = \sigma^2_{Y|X,C} + \frac{\beta^2_{Y|X(C)}\ \sigma^2_{H|X}\ \sigma^2_{X|C}}{\beta^2_{H|X}\ \sigma^2_{X|C} + \sigma^2_{H|X}}$$

$$= \sigma^2_{Y|X,C} + \beta^2_{Y|X(C)} \left(\frac{\sigma^2_{H|X}\ \sigma^2_{X|C}}{\sigma^2_{H|C}} \right) \quad .$$

The equations above provide formulas for the slope parameters and the variance of the model. The remaining model parameter to be considered is the intercept, $\alpha_{Y|H,C}$, which can be expressed as a function of the slope parameters derived above and the mean of the variables Y, H, and C. The formula for the model's intercept is as follows:

$$\alpha_{Y|H,C} = \mu_Y - \left[\left(\beta_{Y|H(C)}\ \mu_H \right) + \left(\beta_{Y|C(H)}\ \mu_C \right) \right] \quad .$$

G10.1.2 MODELING BLOOD-LEAD AS A FUNCTION OF ONE OR MORE VARIABLES MEASURED WITH ERROR AND OTHER SELECT COVARIATES.

Definitions and Assumptions

In the notation that follows, matrices are indicated by bold capital letters and vectors are indicated by underlined letters. Also, squares and square roots of the elements of diagonal matrices are written as the matrix raised to a power(e.g., ^{2} or $^{1/2}$).

Define the following variables:

Y = The response variable, log of blood-lead concentration.

$\underline{R}$ = A vector (or scalar) of observed Rochester Lead-in-Dust Study covariates measured with error (for the Empirical model this vector consists of the log of the area weighted arithmetic mean of floor wipe dust lead loading, the log of the area weighted arithmetic mean of window sill wipe dust lead loading, and the log of the drip-line soil lead concentration).

$\underline{H}$ = A vector (or scalar) of observed HUD National Survey covariates measured with error (for the Empirical model this vector consists of the log of the area weighted arithmetic mean of floor blue nozzle vacuum dust lead loading, the log of the area weighted arithmetic mean of window sill blue nozzle vacuum dust lead loading, and the log of the average soil lead concentration).

$\underline{X}$ = A vector (or scalar) of unobserved covariates measured without error (for the Empirical model this vector consists of the "true" unobserved log of the area weighted arithmetic mean of floor wipe dust lead loading, the "true" unobserved log of the area weighted arithmetic mean of window sill wipe dust lead loading, and the "true" unobserved log of the drip-line soil lead concentration).

$\underline{C}$ = A vector (or scalar) of remaining covariates (for the Empirical model this variable is the scalar paint/pica hazard). These covariates are assumed to be measured using identical methods in the Rochester Lead-in-Dust Study and the HUD National Survey.

The model assumes

(A)

$$\begin{bmatrix} Y \\ \underset{p_1 x 1}{\underline{R}} \\ \underset{p_1 x 1}{\underline{H}} \\ \underset{p_1 x 1}{\underline{X}} \\ \underset{p_2 x 1}{\underline{C}} \end{bmatrix} \sim N \left\{ \begin{bmatrix} \mu_Y \\ \underline{\mu}_R \\ \underline{\mu}_H \\ \underline{\mu}_X \\ \underline{\mu}_C \end{bmatrix}, \begin{bmatrix} \sigma_Y^2 & & & & \\ \underline{\sigma}_{YR} & \underline{\underline{\Sigma}}_{RR} & & & \\ \underline{\sigma}_{YH} & \underline{\underline{\Sigma}}_{RH} & \underline{\underline{\Sigma}}_{HH} & & \\ \underline{\sigma}_{YX} & \underline{\underline{\Sigma}}_{RX} & \underline{\underline{\Sigma}}_{HX} & \underline{\underline{\Sigma}}_{XX} & \\ \underline{\sigma}_{YC} & \underline{\underline{\Sigma}}_{RC} & \underline{\underline{\Sigma}}_{HC} & \underline{\underline{\Sigma}}_{XC} & \underline{\underline{\Sigma}}_{CC} \end{bmatrix} \right\},$$

(B)

For a random sample of size N generated from the distribution in (A):

$$\underset{Nx1}{\underline{Y}} = \underline{1}_N \alpha_{Y|X,C} + \boldsymbol{X} \underline{\beta}_{Y|X(C)} + \boldsymbol{C} \underline{\beta}_{Y|C(X)} + \underline{e}_{Y|X,C} \qquad \underline{e}_{Y|X,C} \sim N\left(\underline{0} , \sigma^2_{Y|X,C} \boldsymbol{I}_N \right)$$

$$\boldsymbol{R}_{Nxp_1} = \boldsymbol{X} + \boldsymbol{E}_{R|X} \qquad \boldsymbol{E}_{R|X} \sim N\left(\boldsymbol{0} , \boldsymbol{I}_N \otimes \underline{\underline{\Sigma}}_{R|X} \right)$$

$$\boldsymbol{H}_{Nxp_1} = \underline{1}_N \underline{\alpha}^T_{H|X} + \boldsymbol{X} \boldsymbol{B}_{H|X} + \boldsymbol{E}_{H|X} \qquad \boldsymbol{E}_{H|X} \sim N\left(\boldsymbol{0} , \boldsymbol{I}_N \otimes \underline{\underline{\Sigma}}_{H|X} \right)$$

$$\boldsymbol{X}_{Nxp_1} = \underline{1}_N \underline{\alpha}^T_{X|C} + \boldsymbol{C} \boldsymbol{B}_{X|C} + \boldsymbol{E}_{X|C} \qquad \boldsymbol{E}_{X|C} \sim N\left(\boldsymbol{0} , \boldsymbol{I}_N \otimes \underline{\underline{\Sigma}}_{X|C} \right) ,$$

and all errors are independent of one another.

The parameters $\alpha_{Y|X,C}$, $\underline{\beta}_{Y|X(C)}$, and $\underline{\beta}_{Y|C(X)}$ represent the intercept and slopes, respectively, associated with a regression of Y on $\underline{X}$(unobserved) and $\underline{C}$; and $\sigma^2_{Y|X,C}$ is the variability in Y unexplained by $\underline{X}$ and $\underline{C}$. $_{R|X}$ is a p_1 by p_1 diagonal matrix with ith diagonal element equal to $\sigma^2_{Ri|Xi}$, the measurement error associated with the ith covariate in Rochester measured with error. $_{H|X}$ is defined analogously for HUD. The ith element of the p_1 by 1 vector $\underline{\alpha}_{H|X}$ represents a location shift in the distribution of the ith variable in $\underline{H}$ relative to the distribution of the ith variable in $\underline{X}$. Similarly, $\mathbf{B}_{H|X}$ is a p_1 by p_1 diagonal matrix with ith diagonal element representing a scale shift in the distribution of the ith variable in $\underline{H}$ relative to the distribution of the ith variable in $\underline{X}$. The p_1 by 1 vector $\underline{\alpha}_{X|C}$ and the p_2 by p_1 matrix $\mathbf{B}_{X|C}$ are the intercepts and slopes, respectively, associated with a regression of $\underline{X}$ on the covariates in $\underline{C}$. $_{X|C}$ is a diagonal matrix with ith diagonal element equal to the variability in the ith element of $\underline{X}$ unexplained by the covariates in $\underline{C}$.

In addition, the calculations that follow rely heavily on the assumption that the conditional distribution of $\underline{X}$ given $\underline{C}$ ($\underline{X}|\underline{C}$) is the same in the Rochester Lead-in-Dust Study and the HUD National Survey. This assumption will be referred to as an assumption of transportability.

Parameter Development

Using assumption (A) of Section G10.2.1,normal distribution theory gives the following result for the conditional distribution of Y given $\underline{H}$ and $\underline{C}$:

$$Y|\underline{H},\underline{C} \sim N(\mu_{Y|H,C}\ ,\ \sigma^2_{Y|H,C})\ ;\ \text{where,}$$

$$\mu_{Y|H,C} = \mu_Y + \begin{bmatrix} \underline{\sigma}^T_{YH} & \underline{\sigma}^T_{YC} \end{bmatrix} \begin{bmatrix} \underline{\underline{\sigma}}_{HH} & \\ \underline{\underline{\sigma}}_{HC} & \underline{\underline{\sigma}}_{CC} \end{bmatrix}^{-1} \begin{bmatrix} \underline{H} - \underline{\mu}_H \\ \underline{C} - \underline{\mu}_C \end{bmatrix}$$

and

$$\sigma^2_{Y|H,C} = \sigma^2_Y - \begin{bmatrix} \underline{\sigma}^T_{YH} & \underline{\sigma}^T_{YC} \end{bmatrix} \begin{bmatrix} \underline{\underline{\sigma}}_{HH} & \\ \underline{\underline{\sigma}}_{HC} & \underline{\underline{\sigma}}_{CC} \end{bmatrix}^{-1} \begin{bmatrix} \underline{\sigma}_{YH} \\ \underline{\sigma}_{YC} \end{bmatrix}.$$

Solving for the inverse above and using (B) for substitutions gives:

$$\mu_{Y|H,C} = \mu_Y + \begin{bmatrix} (\boldsymbol{B}^2_{H|X}\ \ _{X|C} + \ \ _{H|X})^{-1}\ (\boldsymbol{B}_{H|X}\ \ _{X|C}\ \underline{\beta}_{Y|X(C)}) \\ \boldsymbol{B}_{X|C}\ (\boldsymbol{B}^2_{H|X}\ \ _{X|C} + \ \ _{H|X})^{-1}\ (\ _{H|X}\ \underline{\beta}_{Y|X(C)}) + \underline{\beta}_{Y|C(X)} \end{bmatrix}^T \begin{bmatrix} \underline{H} - \underline{\mu}_H \\ \underline{C} - \underline{\mu}_C \end{bmatrix}$$

and

$$\sigma^2_{Y|H,C} = \sigma^2_{Y|X,C} + \underline{\beta}^T_{Y|X(C)}\ (\boldsymbol{B}^2_{H|X}\ \ _{X|C} + \ \ _{H|X})^{-1}\ \ _{H|X}\ \ _{X|C}\ \underline{\beta}_{Y|X(C)}\ .$$

Upon substituting the equality $\Sigma_{H|C} = \boldsymbol{B}^2_{H|X} \Sigma_{X|C} + \Sigma_{H|X}$ into the equation above, assumptions (A) and (B) yield the following slope and variance estimates for the Empirical model:

$$\underline{\beta}_{Y|H(C)} = (\boldsymbol{I}_{p_1} - \Sigma_{H|X} \Sigma^{-1}_{H|C}) \boldsymbol{B}^{-1}_{H|X} \underline{\beta}_{Y|X(C)} ,$$

$$\underline{\beta}_{Y|C(H)} = \underline{\beta}_{Y|C(X)} + \boldsymbol{B}_{X|C} \Sigma_{H|X} \Sigma^{-1}_{H|C} \underline{\beta}_{Y|X(C)} ,$$

and

$$\sigma^2_{Y|H,C} = \sigma^2_{Y|X,C} + \underline{\beta}^T_{Y|X(C)} \Sigma_{H|X} \Sigma^{-1}_{H|C} \Sigma_{X|C} \underline{\beta}_{Y|X(C)} .$$

Finally, the formula used to estimate the Empirical model's intercept is given by:

$$\alpha_{Y|H,C} = \mu_Y - \underline{\beta}^T_{Y|H(C)} \underline{\mu}_H - \underline{\beta}^T_{Y|C(H)} \underline{\mu}_C .$$

G10.1.3 PARAMETER ESTIMATION

Sections G10.1 and G10.2 above provide equations for the model parameters after adjusting for differences between sample collection methods in the Rochester Lead-in-Dust Study and the HUD National Survey. Each variable appearing in the above equations first must be estimated in order to obtain the final estimates of the Empirical model parameters. The following text describes the methodology used to estimate the variables that appear in the final Empirical model formulas of Section G10.2.2. Note that all the variance components described below are provided in Table G10.1.

In the discussion that follows, all estimates for parameters from the HUD National Survey are weighted estimates. The weights correspond to the 1993 American Housing Survey adjusted to 1997. Weights are used because the HUD National Survey is designed to be nationally representative and each observation in HUD is weighted with respect to the population it represents.

Estimation of Parameters Used in Deriving the Empirical Model Slopes and Variance

For estimating the parameters $\underline{\beta}_{Y|X(C)}$, $\underline{\beta}_{Y|C(X)}$, and $\sigma^2_{Y|X,C}$, a classic errors-in-variables model is applied to the Rochester data. The application of this model requires an estimate of the true measurement errors associated with the elements of $\underline{R}$(i.e., $\Sigma_{R|X}$). For further detail on the errors-in-variables model and the estimation of measurement errors associated with both $\underline{R}$ and $\underline{H}$ (i.e., $\Sigma_{R|X}$ and $\Sigma_{H|X}$), see Sections 2 and 3 below.

The ith diagonal element of $\Sigma_{R|C}$ and $\Sigma_{H|C}$ is estimated by the mean squared error from a least squares regression of the ith element of $\underline{R}$ on the covariate vector $\underline{C}$ in Rochester and the

mean squared error from a weighted least squares regression of the ith element of $\underline{H}$ on the covariate vector $\underline{C}$ in HUD, respectively. For example, in the Empirical model, the first diagonal element of $_{R|C}$ is estimated by the mean squared error from the least squares regression of log floor wipe dust lead loading on paint/pica hazard in the Rochester data set.

Using assumption (B) from Section 1.2.1 along with the assumption that all errors are independently distributed yields

$$_{H|C} = \boldsymbol{B}^2_{H|X} \; _{X|C} + \; _{H|X} \text{ and } \; _{R|C} = \; _{X|C} + \; _{R|X} .$$

The estimate of $_{X|C}$ is derived easily from the second equality given above. Using both equalities above, $\mathbf{B}_{H|X}$ is estimated as

$$\boldsymbol{B}_{H|X} = [(\; _{H|C} - \; _{H|X})(\; _{R|C} - \; _{R|X})^{-1}]^{1/2}.$$

Since $\underline{X}$ is a latent variable, the parameter $\mathbf{B}_{X|C}$ cannot be observed. However, from assumption (A) in Section 1.2.1,

$$E(\boldsymbol{R}) = E(\boldsymbol{X}) = \underline{1} \; \underline{\alpha}^T_{X|C} + \boldsymbol{C} \; \boldsymbol{B}_{X|C} .$$

So, $\mathbf{B}_{X|C}$ is estimated by $\mathbf{B}_{R|C}$, which is obtained from a least squares regression of $\underline{R}$ on the covariate vector $\underline{C}$ in Rochester.

Estimation of Parameters Used in Deriving the Empirical Model Intercept

Estimates of the slope parameters, $\underline{\beta}_{Y|H(C)}$ and $\underline{\beta}_{Y|C(H)}$, follow from Section G1-1.3.1 above. The mean parameters, $\underline{\mu}_H$ and $\underline{\mu}_C$, are estimated by weighted means of $\underline{H}$ and $\underline{C}$, respectively, as observed in the HUD National Survey. Unfortunately, Y is not measured in the HUD National Survey; therefore, μ_Y can not be estimated directly from HUD data. As a result, using the intercept formula given in Section G10.1.2 requires an alternative estimate of μ_Y.

Given the intent of the Empirical model, the alternative estimate that is used for μ_Y is the mean of the log of blood-lead concentration in the NHANES III data set. This decision was arrived at for the following reasons:

(1) NHANES III data provide a perfectly legitimate estimate of μ_Y, the national mean of log(blood-lead concentration), with the added appeal of guaranteeing the model's predicted national mean equals the targeted national mean.

(2) The only other sensible estimate of μ_Y, the sample mean from the Rochester study, may be a poor estimator since the distribution of covariates in Rochester is different from the distribution of covariates in HUD. Subsequently, mean blood-lead

concentrations(being a function of the covariates) can be expected to differ across studies as well.

G10.2 REGRESSION PARAMETER ESTIMATION IN THE PRESENCE OF MEASUREMENT ERROR

Let

$$Y = X\beta + \qquad (1)$$

where,

Y = an nx1 vector containing the n values of the dependent variable,

X = an nxp matrix where each column contains the n values of one independent variable in the regression model (in a model with an intercept term, one of the columns would be a column of ones),

β = a px1 vector of regression coefficients, and

= an nx1 vector of random error terms.

In a standard regression model it is assumed that X is a matrix of fixed and known constants, β is a vector of fixed and unknown constants, and is distributed as MVN(0,σ^2I) where MVN(μ,) represents a multivariate normal distribution with mean vector μ and covariance matrix . Estimates of regression parameters for this standard regression model are obtained as follows:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$\hat{\sigma}^2 = Y^T[I - X(X^T X)^{-1} X^T]Y / (n-p) \qquad (2)$$

$$\hat{Cov}(\hat{\beta}) = \hat{\sigma}^2 (X^T X)^{-1}$$

In the presence of measurement error, it is assumed that

$$Y = R\beta + \qquad (3)$$

where,

X = an nxp matrix of fixed but unknown constants representing the values of the independent variables if measured without error;

R = X + is an nxp matrix representing the values of the independent variables observed with measurement error, and

= an nxp matrix of the random measurement errors associated with each of the observed values of the independent variables.

Y and are as defined above. It is assumed that is distributed as MVN(0,I⊗) where is known and is stochastically independent of . Under this measurement error model, estimates of regression parameters are obtained as follows:

$$\hat{\beta} = (R^T R - n\ )^{-1} R^T Y$$

$$MSE_{Y|R} = Y^T[I - R(R^T R)^{-1} R^T]Y / (n-p) \qquad (4)$$

$$\hat{Cov}\ (\hat{\beta}) = MSE_{Y|R}\ (R^T R - n\ )^{-1} R^T R\ (R^T R - n\ )^{-1}$$

These estimators are equivalent to those recommended in Equations (2.2.11) and (2.2.12) by Fuller (Measurement Error Models, 1987).

It can be shown that

1a. The difference between [(R^T R - n) / n] and [X^T X / n] converges in probability to zero as n→∞;

1b. The difference between [(R^T R - (n-p)) / (n-p)] and [X^T X / (n-p)] converges in probability to zero as n→∞; and

2. The difference between [R^T Y / n] and [X^T Y / n] converges in probability to zero as n→∞.

Additionally, it is assumed that

3. X is distributed as MVN(1 μ_x^T,I⊗ $_x$) and is stochastically independent of both and ,

and hence all inferences are based on the conditional distribution of Y given X.

G10.3 DETAILS ON MEASUREMENT ERROR ESTIMATION

The statistical models that account for differences in sample collection methods used in the Rochester Lead-in-Dust Study and the HUD National Survey require estimates of variance components associated with the dust-lead and soil-lead predictor variables in each study. In the notation that follows, a subscript of "f" represents floor dust lead loadings, a subscript of "w" represents window sill dust lead loadings, and a subscript of "s" represents soil lead concentrations. Specifically, we need to obtain the following estimates:

1. The "between homes" variance of observed values of the dust-lead and soil-lead predictor variables for Rochester (σ^2_{Rf}, σ^2_{Rw}, and σ^2_{Rs} corresponding to the diagonal elements of $_{RR}$ from Section G10.1.2 above) and for HUD (σ^2_{Hf}, σ^2_{Hw}, and σ^2_{Hs} corresponding to the diagonal elements of $_{HH}$ from Section G10.1.2 above),

2. After adjusting for the effects of covariates included in the Empirical model, the "between homes" variance of observed values of the dust-lead and soil-lead predictor variables for Rochester ($\sigma^2_{Rf|C}$, $\sigma^2_{Rw|C}$, and $\sigma^2_{Rs|C}$ corresponding to the diagonal elements of $_{R|C}$ from Section G10.1.3 above) and for HUD ($\sigma^2_{Hf|C}$, $\sigma^2_{Hw|C}$, and $\sigma^2_{Hs|C}$ corresponding to the diagonal elements of $_{H|C}$ from Section G10.1.3 above), and

3. The "within homes" variance attributable to measurement error associated with the dust-lead and soil-lead predictor variables for Rochester ($\sigma^2_{Rf|Xf}$, $\sigma^2_{Rw|Xw}$, and $\sigma^2_{Rs|Xs}$ corresponding to the diagonal elements of $_{R|X}$ from Section G10.1.2 above) and for HUD ($\sigma^2_{Hf|Xf}$, $\sigma^2_{Hw|Xw}$, and $\sigma^2_{Hs|Xs}$ corresponding to the diagonal elements of $_{H|X}$ from Section G10.1.2 above).

The following four sections provide details on the methods used to estimate each of the above variance components.

G10.3.1 "BETWEEN HOMES" VARIANCE OF DUST-LEAD AND SOIL-LEAD PREDICTOR VARIABLES

Between home variances of log(floor wipe dust-lead loading), log(window sill wipe dust-lead loading), and log(drip-line soil lead concentration) from the Rochester Lead-in-Dust Study are represented by σ^2_{Rf}, σ^2_{Rw}, and σ^2_{Rs}, respectively. Each variance is estimated by the sample variance of the respective variable as observed in the Rochester dataset. Specifically,

$$\sigma^2_{Rf} = \frac{1}{N-1}\sum_{i=1}^{N}(Rf_i - \bar{R}f)^2 \ , \ \sigma^2_{Rw} = \frac{1}{N-1}\sum_{i=1}^{N}(Rw_i - \bar{R}w)^2 \ , \text{ and } \sigma^2_{Rs} = \frac{1}{N-1}\sum_{i=1}^{N}(Rs_i - \bar{R}s)^2 \ ,$$

where Rf_i and Rw_i represent the floor and window sill dust-lead predictor variables and Rs_i represents the soil-lead predictor variable associated with each home in the Rochester Study. $\bar{R}f$ represents the sample mean of log(floor dust lead loading) among all homes from the Rochester Study, $\bar{R}w$ represents the sample mean of log(window sill dust lead loading) among all homes from the Rochester Study, and $\bar{R}s$ represents the sample mean of log(drip-line soil lead concentration) among all homes from the Rochester Study.

Between home variances of log(blue nozzle floor dust lead loading), log(blue nozzle window sill dust lead loading), and log(average soil lead concentration) from the HUD Survey are represented by σ^2_{Hf}, σ^2_{Hw}, and σ^2_{Hs}, respectively. In contrast to the Rochester between home variance estimates described above, HUD Survey between home variance estimates are weighted.

Each observation is weighted using weights from the 1993 American Housing Survey adjusted to 1997. Weights are used because the HUD Survey is designed to be nationally representative and each observation in HUD is weighted with respect to the population it represents. Specifically,

$$\sigma^2_{Hf} = \frac{\sum_{i=1}^{N} w_i (Hf_i - \bar{H}f)^2}{\sum_{i=1}^{N} w_i - 1}, \quad \sigma^2_{Hw} = \frac{\sum_{i=1}^{N} w_i (Hw_i - \bar{H}w)^2}{\sum_{i=1}^{N} w_i - 1}, \text{ and } \sigma^2_{Hs} = \frac{\sum_{i=1}^{N} w_i (Hs_i - \bar{H}s)^2}{\sum_{i=1}^{N} w_i - 1},$$

where w_i represents the population weight for the ith home in the HUD National Survey, Hf_i, Hw_i, and Hs_i represent the floor and window sill dust-lead predictor variables and soil-lead predictor variable associated with each house in the HUD National Survey, $\sum_{i=1}^{N} w_i = N$, and $\bar{H}f$, $\bar{H}w$, and $\bar{H}s$ are weighted means calculated as follows:

$$\bar{H}f = \frac{\sum_{i=1}^{N} w_i Hf_i}{\sum_{i=1}^{N} w_i}, \quad \bar{H}w = \frac{\sum_{i=1}^{N} w_i Hw_i}{\sum_{i=1}^{N} w_i}, \text{ and } \bar{H}s = \frac{\sum_{i=1}^{N} w_i Hs_i}{\sum_{i=1}^{N} w_i}.$$

G10.3.2 COVARIATE ADJUSTED "BETWEEN HOMES" VARIANCE OF DUST-LEAD AND SOIL-LEAD PREDICTOR VARIABLES

$\sigma^2_{Rf|C}$, $\sigma^2_{Rw|C}$, $\sigma^2_{Rs|C}$, $\sigma^2_{Hf|C}$, $\sigma^2_{Hw|C}$, and $\sigma^2_{Hs|C}$ represent the portion of between home variance (σ^2_{Rf}, σ^2_{Rw}, σ^2_{Rs}, σ^2_{Hf}, σ^2_{Hw}, and σ^2_{Hs}, respectively) that remains after adjusting for the other covariates included in the Empirical model. An estimate of these quantities can be obtained from the mean squared error of a least squares regression of the variables (Rf, Rw, Rs, Hf, Hw, or Hs) on the other covariates in the Empirical model. The least squares regression model treats the covariates as fixed; and the resulting mean squared error estimates the remaining variability of the variable in the presence of the fixed covariates.

The covariate adjusted between home variances from the Rochester Lead-in-Dust Study, $\sigma^2_{Rf|C}$, $\sigma^2_{Rw|C}$, and $\sigma^2_{Rs|C}$, are estimated using mean squared errors obtained from ordinary least squares regressions of log(floor wipe dust lead loading), log(window sill wipe dust lead loading), and log(drip-line soil lead concentration) on the remaining model covariates, respectively. Similarly, the covariate adjusted between home variances from the HUD Survey, $\sigma^2_{Hf|C}$, $\sigma^2_{Hw|C}$, and $\sigma^2_{Hs|C}$, are estimated using mean squared errors obtained from weighted least squares regressions of log(floor wipe dust lead loading), log(window sill wipe dust lead loading), and log(average soil lead concentration) on the remaining model covariates, respectively. Again, least squares regressions involving HUD data are weighted because the HUD Survey is designed to be nationally representative.

G10.3.3 MEASUREMENT ERROR ASSOCIATED WITH PREDICTOR VARIABLES

The dust-lead predictor variables in the statistical models represent area-weighted arithmetic average individual sample dust-lead loadings from floors and window sills. The following equation represents the three sources of variability that must be accounted for in an estimate of measurement error for these dust-lead predictor variables:

$$\sigma^2_{\text{Measurement Error}} = \sigma^2_{\text{Spatial}} + \sigma^2_{\text{Sampling}} + \sigma^2_{\text{Laboratory}} ,$$

where $\sigma^2_{\text{Spatial}}$ represents the variability in dust-lead levels among all possible locations on the surface being tested, $\sigma^2_{\text{Sampling}}$ represents variability in the collection of dust from the surface, and $\sigma^2_{\text{Laboratory}}$ represents variability in the chemical analysis of the sample. This definition of measurement error is consistent with the interpretation of each predictor variable as exposure to lead from floor or window sill dust found at the primary residence at the time of sampling. Thus there was no attempt to estimate a component of variation associated with temporal variability. The following two subsections contain details on estimating the measurement associated with dust-lead and soil-lead predictor variables.

G10.3.3.1 Measurement Error Associated with Dust-Lead Predictor Variables

Several sources of data were considered for providing information about the variability in dust sample results due to measurement error, including field duplicate data and data that included multiple dust samples (of a given component type) collected from within the same house. Since the predictor variables included in the statistical models represented area weighted averages of multiple dust sample results collected within a house, the individual sample lead loading results from the Rochester Lead-in-Dust Study and the HUD National Survey were used to assess the measurement error. Specifically, let

$Dust_{ijk}$ represent the dust-lead loading from the kth component type (floor or window sill) from the jth location within the ith residential unit,

$Area_{ijk}$ represent the area of the sample from the kth component type from the jth location within the ith residential unit, and

The following model was then fitted separately for floors and window sills from each study to estimate the within house variability in dust-lead loadings between individual dust samples:

$$\ln(Dust_{ijk}) = \ln(\mu_k) + H_{ik} + E_{ijk} ,$$

where μ_k is the geometric mean of $Dust_{ijk}$ among all samples of component k, H_{ik} is the random effect associated with the ith House, and E_{ijk} is the random within-house error term associated with $Dust_{ijk}$. H_{ik} is assumed to follow a normal distribution with mean zero and variance $\sigma^2_{\text{Between Houses}}$, and E_{ijk} is assumed to follow a normal distribution with mean zero and variance $\sigma^2_{\text{Within Houses}}$.

$\sigma^2_{\text{Between Houses}}$ characterizes the variability between houses. $\sigma^2_{\text{Within Houses}}$ characterizes the variability within a house; attributed to a combination of spatial, sampling, and laboratory variability. The following two subsections describe how weights were used with the above model to calculate the measurement error variance components $\sigma^2_{\text{Rf|Xf}}$ and $\sigma^2_{\text{Rw|Xw}}$ corresponding to the Rochester Lead-in-Dust Study, and $\sigma^2_{\text{Hf|Xf}}$ and $\sigma^2_{\text{Hw|Xw}}$ corresponding to the HUD National Survey.

Rochester Lead-in-Dust Study

Since area weighted (arithmetic) mean floor and window sill dust-lead loadings were used to characterize the dust-lead levels in each house in the Rochester Lead-in-Dust Study, the above model was fitted using weights corresponding to the percent of total area that was associated with each sample:

$$Weight_{ijk} = \frac{Area_{ijk}}{\sum_{j=1}^{N_{ik}} Area_{ijk}}$$

where n_{ik} is the number of samples collected from component k within the ith house.

Values of $\sigma^2_{\text{Within Houses}}$ calculated in this weighted analysis are used as estimates of $\sigma^2_{\text{Rf|Xf}}$ and $\sigma^2_{\text{Rw|Xw}}$ in the statistical models described in Sections 1 and 2 of this appendix. In actuality, these estimates of $\sigma^2_{\text{Rf|Xf}}$ and $\sigma^2_{\text{Rw|Xw}}$ correspond more closely to measurement error in area weighted geometric mean dust-lead loadings from floors and window sills within each house. Table G10-1 provides estimates of $\sigma^2_{\text{Rf|Xf}}$ and $\sigma^2_{\text{Rw|Xw}}$ as calculated from the Rochester Lead-in-Dust Study data.

HUD National Survey

Since area weighted (arithmetic) mean floor and window sill dust-lead loadings were used to characterize the dust-lead levels in each house in the HUD National Survey, the above model was fitted using a combination of weights corresponding to the percent of total area that was associated with each sample, and the survey weight associated with each home sampled:

$$Weight_{ijk} = \frac{Area_{ijk}}{\sum_{j=1}^{N_{ik}} Area_{ijk}} \cdot \frac{HSW_i}{\frac{1}{N}\sum_{i=1}^{n} HSW_i}$$

where n_{ik} is the number of samples collected from component k within the ith house, n is the number of homes included in the HUD National Survey, and HSW_i is the survey weight associated with the ith home in the HUD National Survey.

Table G10-1. Components of Variation Used to Implement an Adjustment of the Rochester Multi-Media Predictive Model for Use with Environmental Lead Levels as Measured in the HUD National Survey.

Study	Parameter	Final Empirical Model

Rochester Lead-in-Dust	$^2_{Rf\|Xf}$	0.2082
	$^2_{Rw\|Xw}$	0.5708
	$^2_{Rs\|Xs}$	0.3898
	$^2_{Rf\|C}$	1.3323
	$^2_{Rw\|C}$	1.8505
	$^2_{Rs\|C}$	1.6497
	$^2_{Rf}$	1.3410
	$^2_{Rw}$	1.8592
	$^2_{Rs}$	1.6640
HUD National Survey	$^2_{Hf\|Xf}$	0.6125
	$^2_{Hw\|Xw}$	1.6937
	$^2_{Hs\|Xs}$	0.3016
	$^2_{Hf\|C}$	2.3589
	$^2_{Hw\|C}$	5.2881
	$^2_{Hs\|C}$	2.2125
	$^2_{Hf}$	2.3767
	$^2_{Hw}$	5.3225
	$^2_{Hs}$	2.2434

Values of $\sigma^2_{\text{Within Houses}}$ calculated in this weighted analysis are used as estimates of $\sigma^2_{Hf|Xf}$ and $\sigma^2_{Hw|Xw}$ in the statistical models described in Sections 1 and 2 of this appendix. In actuality, these estimates of $\sigma^2_{Hf|Xf}$ and $\sigma^2_{Hw|Xw}$ correspond more closely to measurement error in area weighted geometric mean dust-lead loadings from floors and window sills within each house. Table G10-1 provides estimates of $\sigma^2_{Hf|Xf}$ and $\sigma^2_{Hw|Xw}$ as calculated from the HUD National Survey data.

G10.3.3.2 Measurement Error Associated with Soil-Lead Predictor Variables

Due to the fact that there was analytical information available from only one composite drip-line soil sample collected from each home in Rochester, we were unable to derive an estimate of measurement error ($\sigma^2_{Rs|Xs}$) using data observed in the Rochester Lead-in-Dust Study. We therefore derived estimates of measurement error associated with soil-lead predictor variables in both the Rochester Study ($\sigma^2_{Rs|Xs}$) and the HUD National Survey ($\sigma^2_{Hs|Xs}$) using data collected in the HUD National Survey.

Up to three different soil samples were collected from each home in the HUD National Survey: an entryway soil sample, a drip-line soil sample, and a remote soil sample. The soil-lead predictor variables used in the Empirical Model can be regarded as weighted averages of these multiple soil sample results collected within each HUD National Survey home. Specifically, we considered the Rochester soil-lead predictor variable to be representative of the average between entryway and drip-line soil samples collected in the HUD National Survey (each sample receiving weight of 0.5). The HUD National Survey predictor variable was constructed as the average between the remote soil sample and the average between entryway and drip-line soil samples (remote sample receiving weight of 0.5, and drip-line and entryway samples each receiving weight of 0.25). The individual soil-lead concentration results from the HUD National Survey were used to assess the measurement error variance components as follows:

Let $Soil_{ij}$ represent the soil-lead concentration from the jth location within the ith residential unit. The following model was then fitted to estimate the within house variability in soil-lead concentration between individual soil samples:

$$\ln(Soil_{ij}) = \ln(\mu) + H_i + E_{ij} ,$$

where μ_k is the geometric mean of $Soil_{ij}$ among all samples, H_i is the random effect associated with the ith House, and E_{ij} is the random within-house error term associated with $Soil_{ij}$. H_i is assumed to follow a normal distribution with mean zero and variance $\sigma^2_{Between\ Houses}$, and E_{ij} is assumed to follow a normal distribution with mean zero and variance $\sigma^2_{Within\ Houses}$.

$\sigma^2_{Between\ Houses}$ characterizes the variability between houses. $\sigma^2_{Within\ Houses}$ characterizes the variability within a house; attributed to a combination of spatial, sampling, and laboratory variability. Weights were used with the above model to calculate the measurement error variance components $\sigma^2_{Rs|Xs}$ corresponding to the Rochester Lead-in-Dust Study, and $\sigma^2_{Hs|Xs}$ corresponding to the HUD National Survey as follows:

$$Weight_{ij} = \frac{W_{ij} \cdot HSW_i}{\frac{1}{N}\sum_{i=1}^{n} HSW_i}$$

where n is the number of homes included in the HUD National Survey, HSW_i is the survey weight associated with the ith home in the HUD National Survey, and W_{ij} is the weight corresponding to each individual sample being averaged:

Soil Sample Location	Value of W_{ij} when Estimating	
	$^2_{Rs|Xs}$	$^2_{Rs|Xs}$

Drip-line	0.5	0.25
Entryway	0.5	0.25
Remote	0.0	0.5

Values of $\sigma^2_{\text{Within Houses}}$ calculated in this weighted analysis are used as estimates of $\sigma^2_{Rs|Xs}$ and $\sigma^2_{Hs|Xs}$ in the statistical models described in Sections 1 and 2 of this appendix. Table G10-1 provides estimates of $\sigma^2_{Rs|Xs}$ and $\sigma^2_{Hs|Xs}$ as calculated from the HUD National Survey data.

G10.3.4 EFFECT OF IMPUTING BLUE NOZZLE WINDOW SILL DUST LEAD LOADINGS IN THE HUD DATASET ON ESTIMATED VARIANCE COMPONENTS

The floor and window sill dust-lead loading predictor variable was imputed for several of the homes in the HUD National Survey (for homes that did not include any dust samples from window sills) in an effort to keep as many homes in the analysis as possible, and thus maintain its property of being nationally representative (with appropriate survey weights).

The HUD sample used for calculating σ^2_{Hf}, σ^2_{Hw}, and σ^2_{Hs} includes imputed values, and is the same for the preliminary and final Empirical models; therefore the estimate of σ^2_{Hf} is consistent across the rows in Table G10-1. The HUD sample used for calculating $\sigma^2_{Hf|C}$, $\sigma^2_{Hw|C}$, and $\sigma^2_{Hs|C}$ also includes imputed values.

In contrast, $\sigma^2_{Hf|Xf}$, $\sigma^2_{Hw|Xw}$, and $\sigma^2_{Hs|Xs}$ can only be estimated using those houses in which floor and window sill dust samples and soil samples were collected. Values for $\sigma^2_{Hf|Xf}$ were therefore calculated separately for each version of the empirical model.

G10.3.5 ESTIMATED COMPONENTS OF VARIATION

The following table provides the components of variation used to implement an adjustment of the Rochester multi-media predictive model for use with environmental lead levels as measured in the HUD National Survey.

G10.4 BOOTSTRAP ESTIMATION OF STANDARD ERRORS

In ordinary least squares regression, formulas are readily available for calculating standard errors associated with the model's parameter estimates. For the parameter estimates of the model that accounts for differences in sample collection methods used in the Rochester Lead-in-Dust Study and the HUD National Survey, no such simple formulas exist. As a result, 48 standard errors can only be approximated. The method of approximation used for estimating the standard errors corresponding to the parameters of the adjusted model is a basic bootstrap algorithm, which is described below. Note that the following definitions and algorithm are taken directly from Efron and Tibshirani, "An Introduction to the Bootstrap," 1993 pp. 45-47.

Let $\vec{x} = (x_1, x_2, ..., x_n)$ represent a sample dataset.

Let $\hat{\theta}(\vec{x})$ be an estimator of a parameter of interest θ, where $\hat{\theta}(\vec{x})$ is such that its standard error is not easily obtained.

Define $\hat{F}$ to be the empirical distribution that assigns probability 1/n to each of the n observations in the sample dataset.

Define a bootstrap sample, $\vec{x}^{(b)}$, as a random sample of size n drawn with replacement from $\hat{F}$.

A bootstrap estimate of the standard error of $\hat{\theta}(\vec{x})$ is obtained as follows:

1. Collect B independent bootstrap samples, $\vec{x}_1^{(b)}, \vec{x}_2^{(b)}, \ldots, \vec{x}_B^{(b)}$.
2. For each bootstrap sample, calculate $\hat{\theta}(\vec{x}_i^{(b)})$, i=1,2,..., B.
3. Estimate the standard error of $\hat{\theta}(\vec{x})$ as:

$$\hat{se}_B = \left\{ \sum_{i=1}^{B} \left[\hat{\theta}(\vec{x}_i^{(b)}) - \bar{\theta}^{(b)} \right] / (B-1) \right\}^{1/2}, \text{ where } \bar{\theta}^{(b)} = \sum_{i=1}^{B} \hat{\theta}(\vec{x}_i^{(b)})/B .$$

The above algorithm is used to estimate the standard errors of the estimators described in Section 1 ($\beta_{Y|Hf(Hw,Hs,C)}$, $\beta_{Y|Hw(Hf,Hs,C)}$, $\beta_{Y|Hs(Hf,Hw,C)}$, $\beta_{Y|C(Hf,Hw,Hs)}$, and $\alpha_{Y|Hf,Hw,Hs,C}$). Because the adjustment procedure is based on data from the Rochester Lead-in-Dust Study, the Rochester dataset is treated as the sample dataset, $\vec{x}$. Data from the HUD National Survey are held fixed in the implementation of the algorithm. In essence, the adjusted model parameters are viewed as functions of sample data(Rochester dataset) that are calibrated to correspond to population values(HUD dataset). Thus, their variability is assumed, in this preliminary assessment, to stem from the Rochester dataset only.

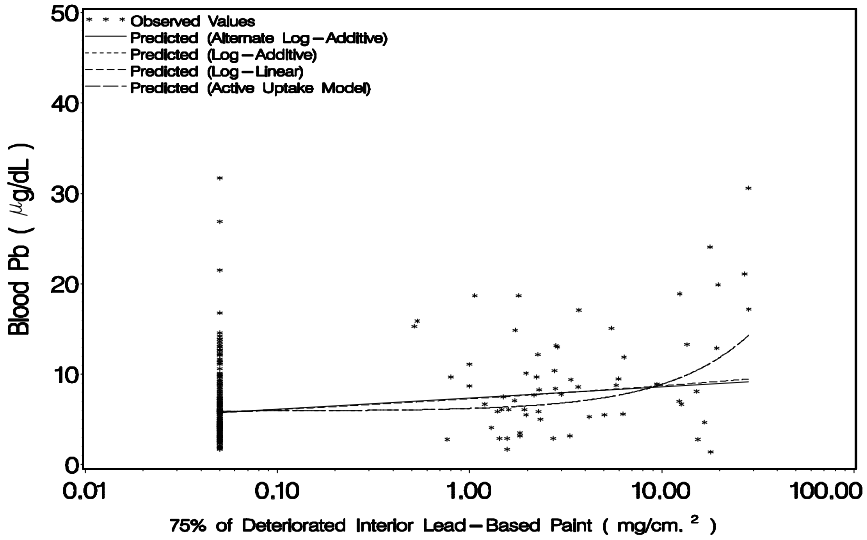
Finally, observe that,

$$\lim_{B \to \infty} \hat{se}_B = se_{\hat{F}} .$$

That is, as the number of bootstrap replications increases, the estimated standard error approaches the population standard error; where the population distribution is estimated by $\hat{F}$. Efron and Tibshirani (1993) recommend between 25 and 200 bootstrap replications for adequate approximations. 200 bootstrap replications were used in the application of the bootstrap algorithm to approximate the standard errors of parameters in the adjusted model.

G11: Appendix on Bivariate Relationships Between Blood Lead and Potential Lead Exposure Predictor Variables

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.99 (0.0545)	0.08 (0.0205)	--	0.3550	0.0668	-183.73	205
Log-Additive	5.93 (0.2722)	0.30 (0.0932)	--	0.3512	0.0767	-182.64	205
Alternate Log-Additive	7.37 (0.4378)	0.53 (0.1573)	--	0.3558	0.0647	-183.95	205
Active Uptake	5.93 (0.2721)	0.30 (0.0932)	4.9E115 (0.000)	0.3547	0.0767	-182.64	205



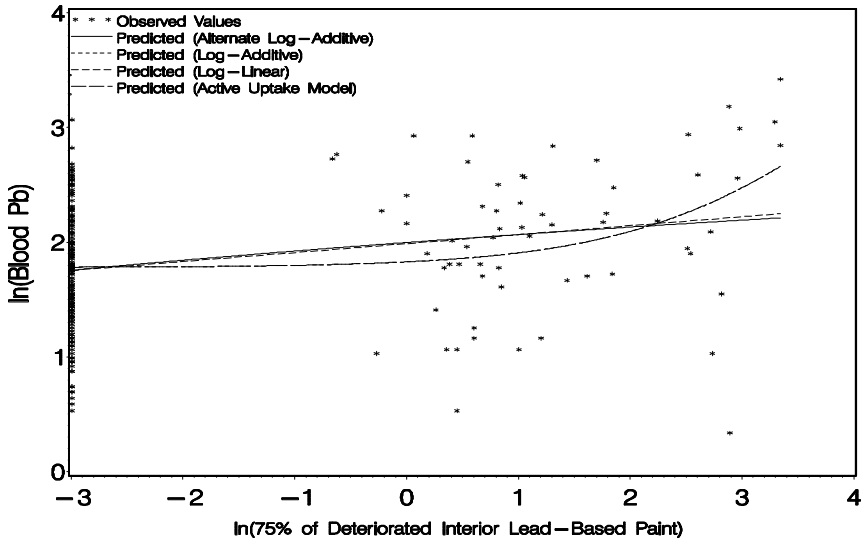


Figure G11-1. Bivariate Relationship Between Blood-Lead Concentration and the 75th Percentile of Deteriorated Interior Lead-Based Paint.

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.89 (0.0459)	0.03 (0.0167)	--	0.3733	0.0187	-188.88	205
Log-Additive	6.04 (0.3015)	0.08 (0.0410)	--	0.3722	0.0215	-188.58	205
Alternate Log-Additive	6.62 (0.3140)	0.21 (0.1125)	--	0.3733	0.0186	-188.89	205
Active Uptake	6.04 (29.1731)	0.08 (0.8766)	6.8E8 (3.7E15)	0.3759	0.0215	-188.59	205

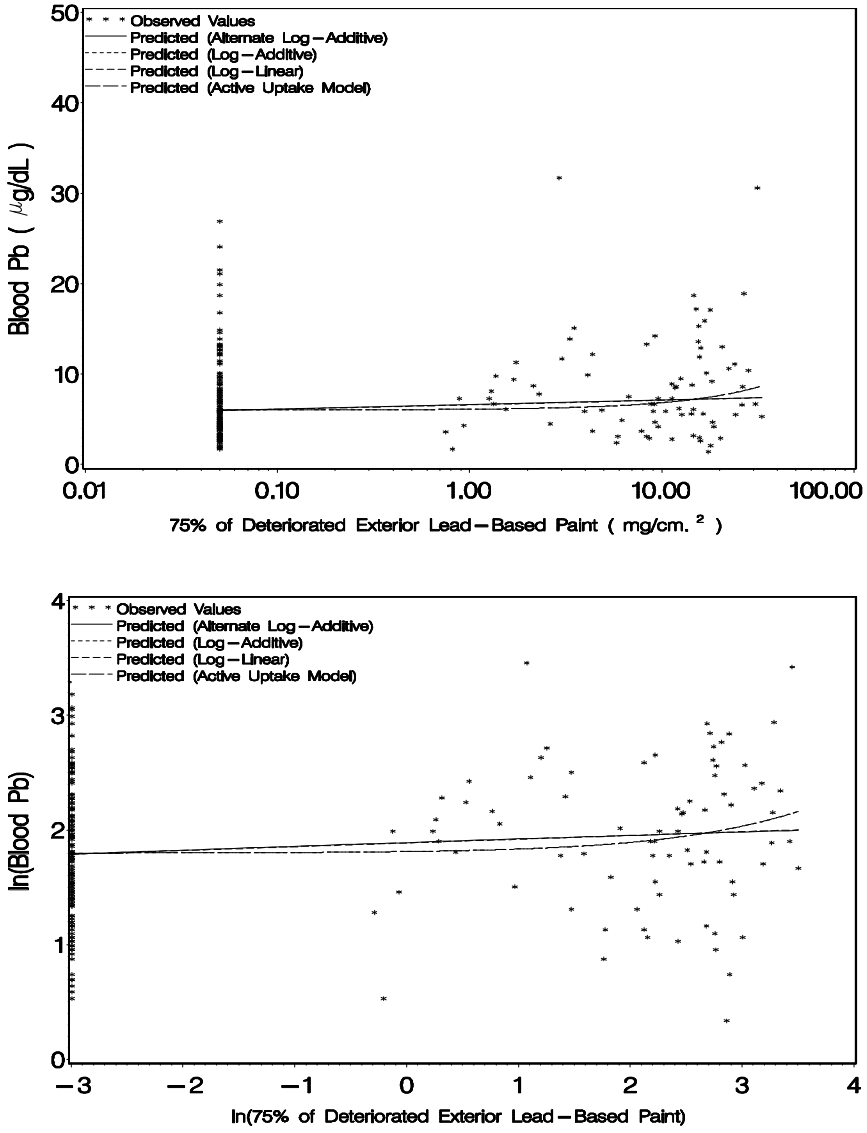


Figure G11-2. Bivariate Relationship Between Blood-Lead Concentration and the 75th Percentile of Deteriorated Exterior Lead-Based Paint.

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.38 (0.1130)	0.17 (0.0365)	--	0.3433	0.0959	-173.22	197
Log-Additive	6.37 (0.2807)	0.00 (0.0006)	--	0.3741	0.0148	-181.68	197
Alternate Log-Additive	3.60 (0.6354)	1.02 (0.2343)	--	0.3452	0.0909	-173.76	197
Active Uptake	6.58 (1.3665)	0.31 (0.1902)	13.22 (3.5208)	0.3407	0.1119	-171.47	197

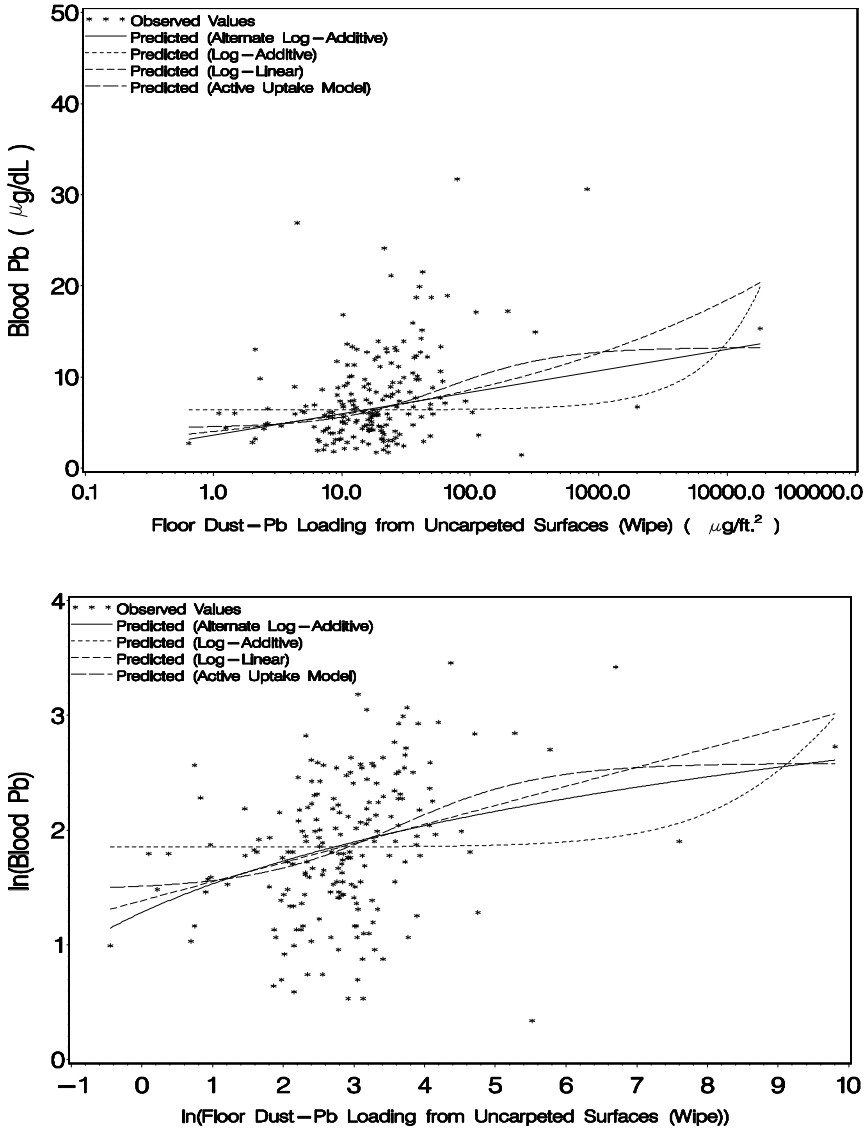


Figure G11-3. Bivariate Relationship Between Blood-Lead Concentration and Floor Dust-Lead Loading from Uncarpeted Surfaces (Wipe Samples).

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.63 (0.0636)	0.12 (0.0254)	--	0.3451	0.0928	-180.83	205
Log-Additive	6.23 (0.2751)	0.01 (0.0031)	--	0.3708	0.0252	-188.20	205
Alternate Log-Additive	5.40 (0.3133)	0.55 (0.1350)	--	0.3515	0.0760	-182.71	205
Active Uptake	7.40 (1.2816)	0.60 (0.3635)	12.68 (2.8613)	0.3351	0.1277	-176.81	205

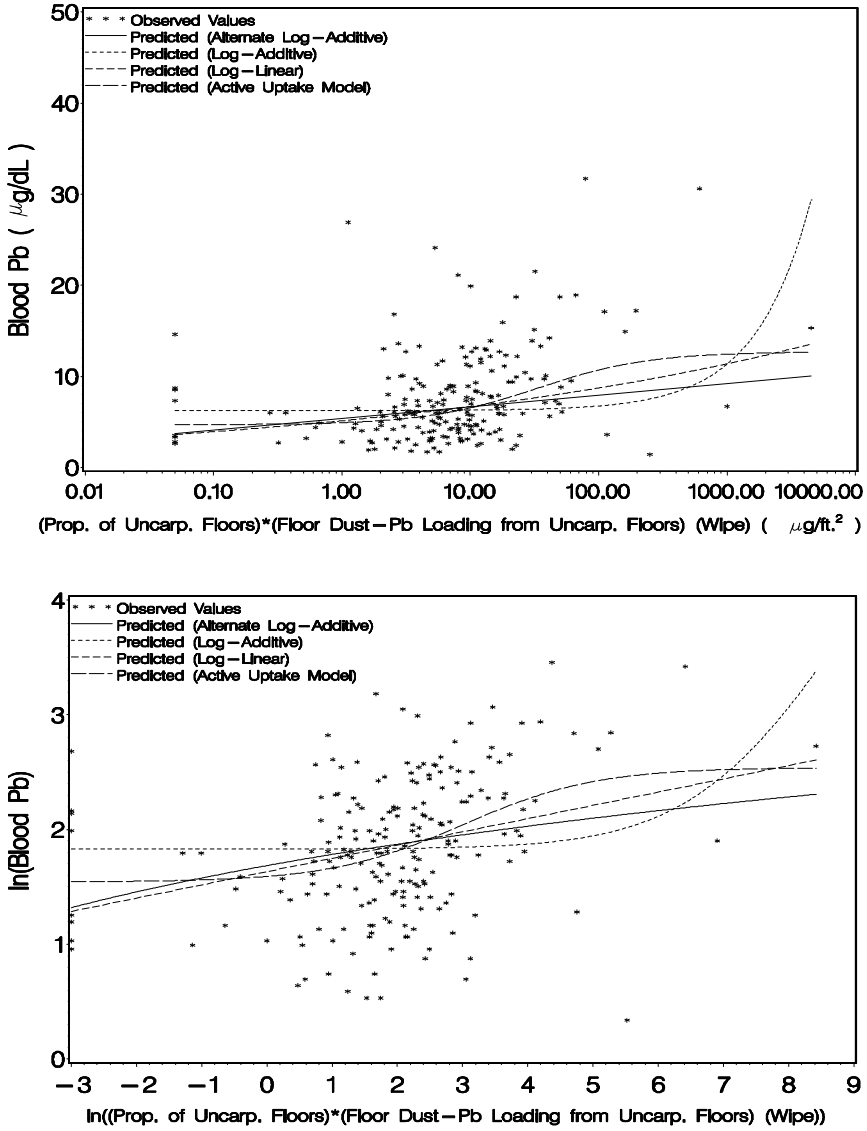


Figure G11-4. Bivariate Relationship Between Blood-Lead Concentration and (Proportion of Uncarpeted Floors Samples)*(Floor Dust-Lead Loading from Uncarpeted Floors) (Wipe Samples).

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.56 (0.1066)	0.10 (0.0385)	--	0.3363	0.0368	-155.47	179
Log-Additive	6.16 (0.2734)	0.00 (0.0002)	--	0.3492	0.00	-158.82	179
Alternate Log-Additive	4.29 (0.6156)	0.76 (0.2481)	--	0.3334	0.0453	-154.67	179
Active Uptake	15.00 (0.0000)	0.00 (0.0012)	10.44 (0.7862)	0.3532	0.00	-158.83	179

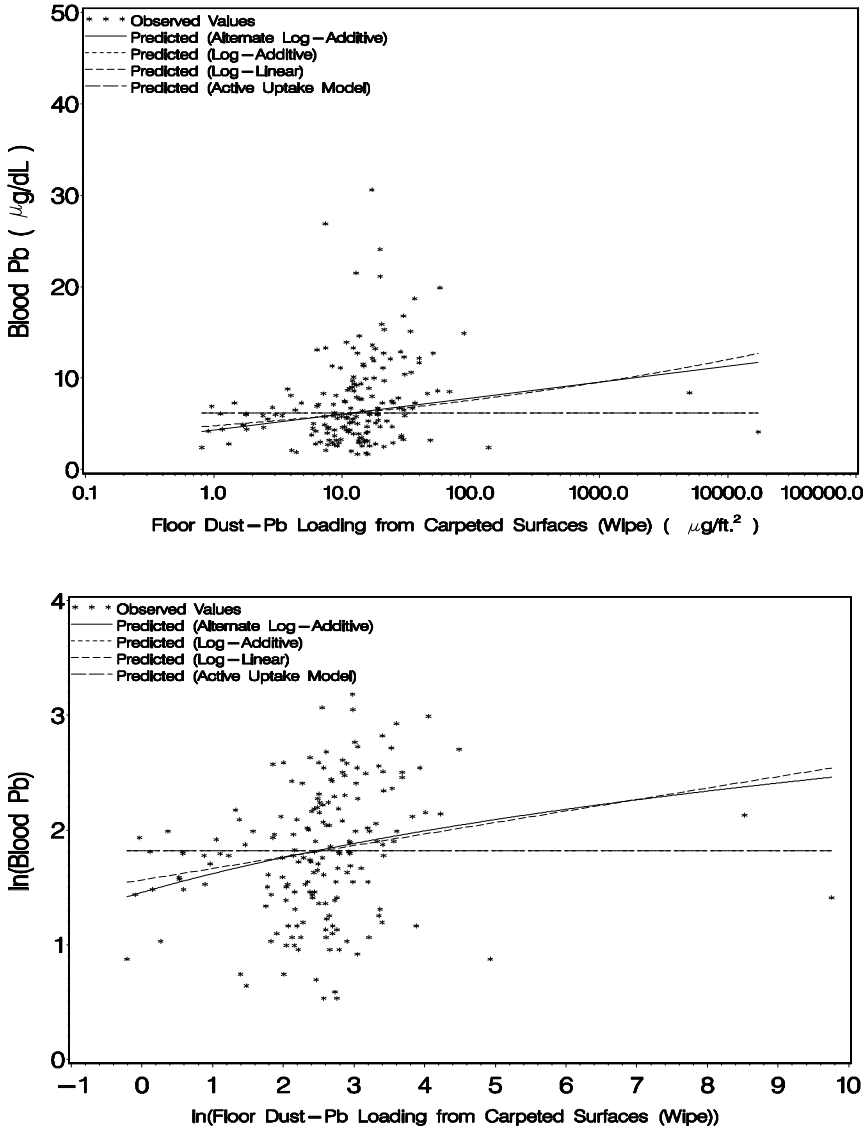


Figure G11-5. Bivariate Relationship Between Blood-Lead Concentration and Floor Dust-Lead Loading from Carpeted Surfaces (Wipe Samples).

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.85 (0.0511)	0.00 (0.0217)	--	0.3804	0.00	-190.81	205
Log-Additive	6.38 (0.2764)	0.00 (0.0004)	--	0.3804	0.00	-190.81	205
Alternate Log-Additive	6.38 (0.3262)	0.00 (0.1384)	--	0.3804	0.00	-190.81	205
Active Uptake	16.22 (0.0000)	0.00 (0.0028)	10.52 (0.7520)	0.3842	0.00	-190.82	205

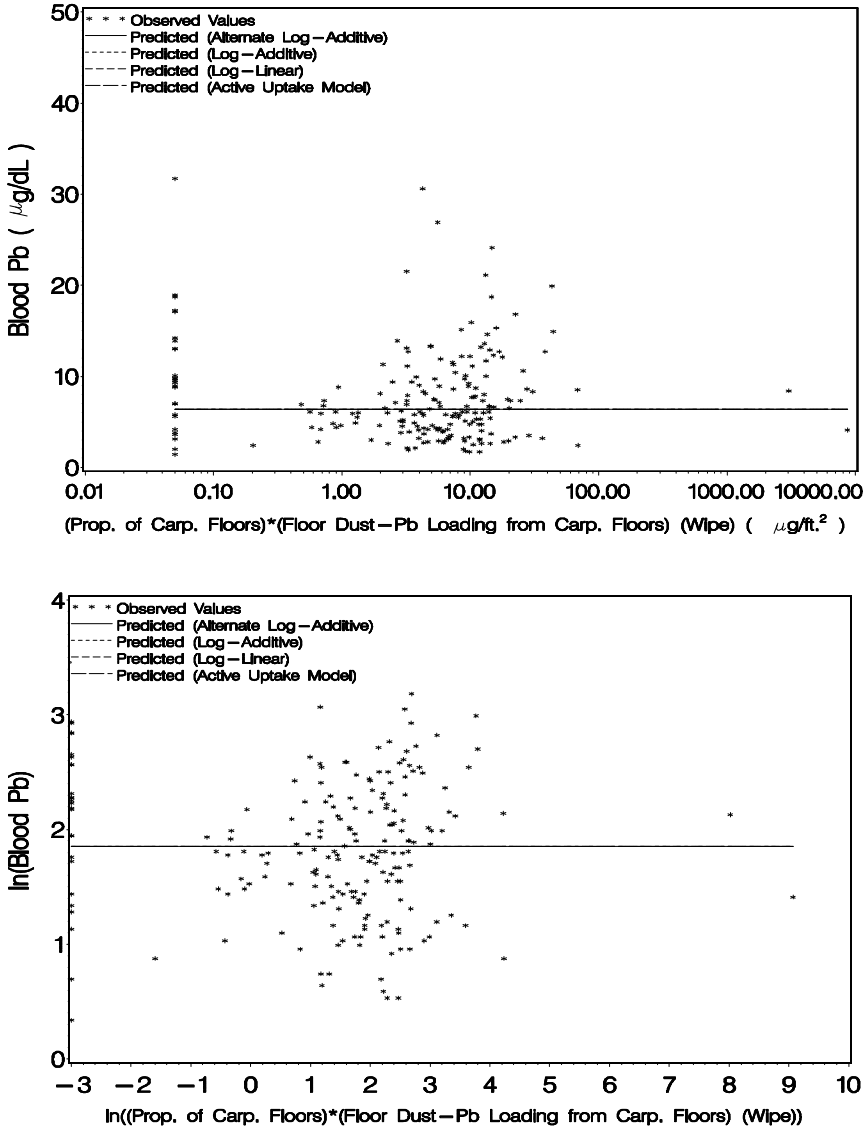


Figure G11-6. Bivariate Relationship Between Blood-Lead Concentration and (Proportion of Carpeted Floors Sampled)*(Floor Dust Lead-Loading from Carpeted Floors (Wipe Samples).

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.04 (0.1658)	0.15 (0.0304)	--	0.3409	0.1144	-171.65	196
Log-Additive	5.58 (0.2956)	0.00 (0.0005)	--	0.3525	0.0845	-174.91	196
Alternate Log-Additive	2.36 (0.8614)	0.77 (0.1731)	--	0.3476	0.0971	-173.56	196
Active Uptake	7.42 (1.2696)	0.02 (0.0144)	12.63 (3.0407)	0.3369	0.1339	-169.48	196

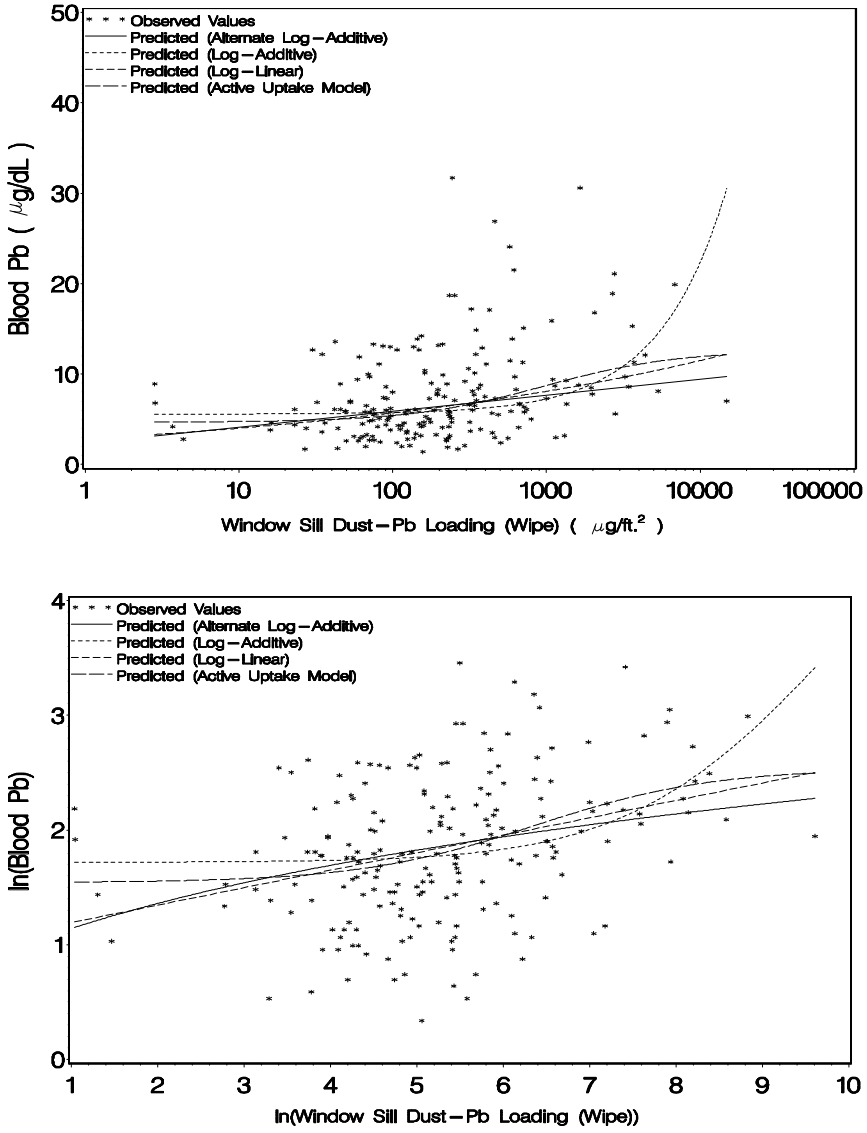


Figure G11-7. Bivariate Relationship Between Blood-Lead Concentration and Window Sill Dust-Lead Loading (Wipe Samples).

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.17 (0.1600)	0.08 (0.0183)	--	0.3533	0.0929	-168.85	189
Log-Additive	5.85 (0.2927)	0.00 (0.0000)	--	0.3676	0.0561	-172.61	189
Alternate Log-Additive	2.45 (0.8881)	0.47 (0.1123)	--	0.3550	0.0884	-169.31	189
Active Uptake	11.69 (2.6346)	0.00 (0.0014)	8.89 (1.1742)	0.3568	0.0938	-168.77	189

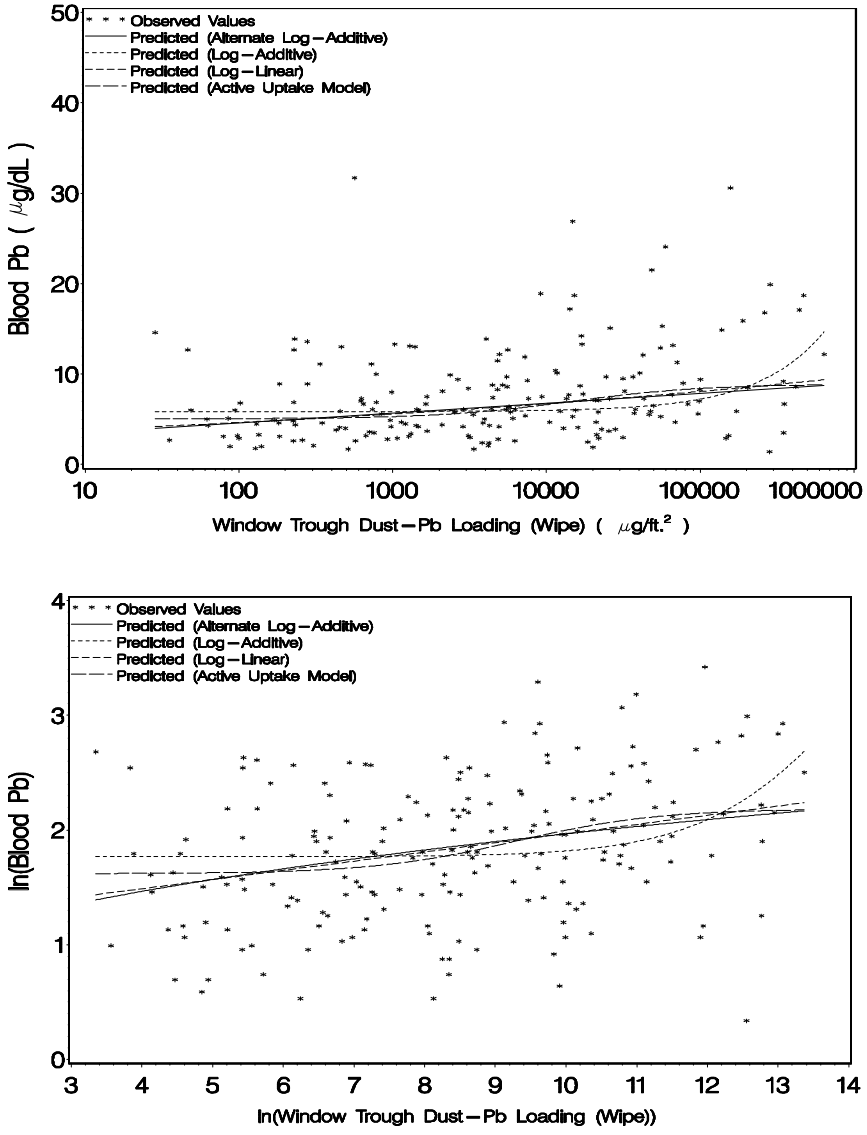
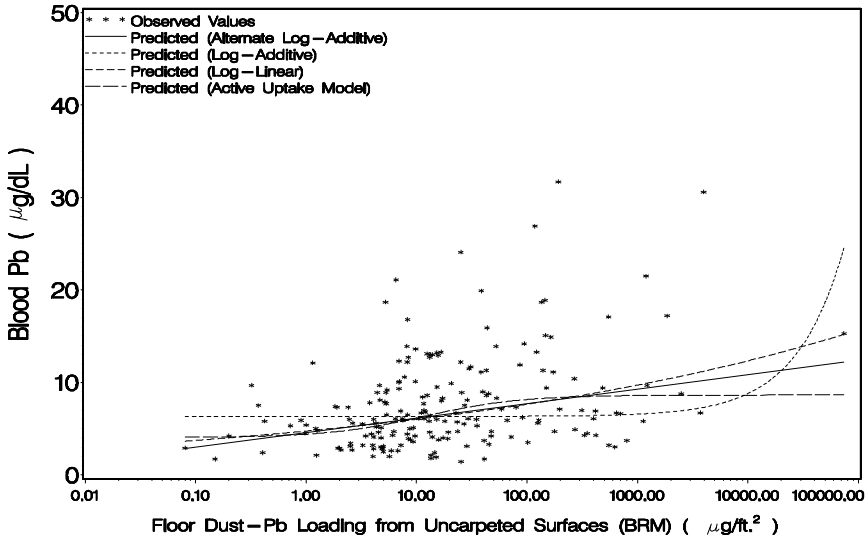


Figure G11-8. Bivariate Relationship Between Blood-Lead Concentration and Window Well Dust-Lead Loading (Wipe Samples).

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.56 (0.0726)	0.10 (0.0205)	--	0.3383	0.1186	-168.27	193
Log-Additive	6.32 (0.2830)	0.00 (0.0002)	--	0.3769	0.0182	-178.68	193
Alternate Log-Additive	4.59 (0.3635)	0.68 (0.1285)	--	0.3375	0.1207	-168.04	193
Active Uptake	7.58 (2.3187)	1.38 (0.8114)	8.63 (0.8053)	0.3417	0.1190	-168.23	193



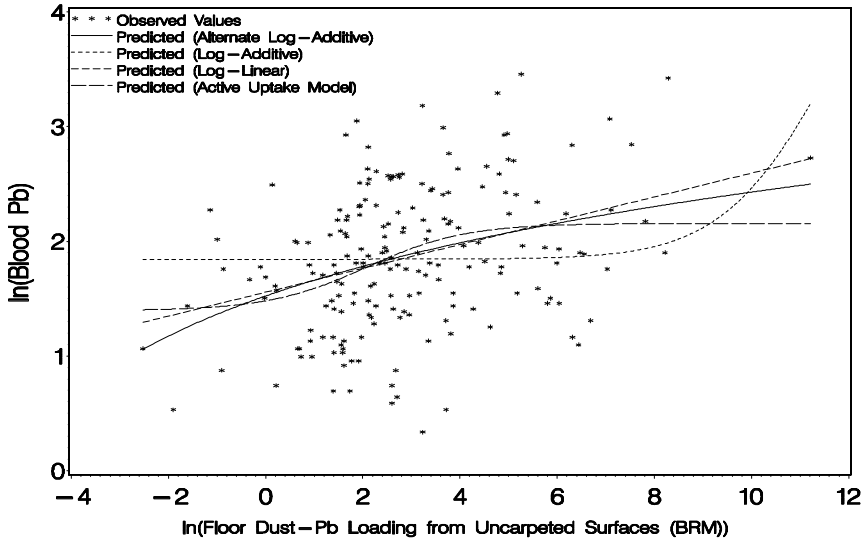
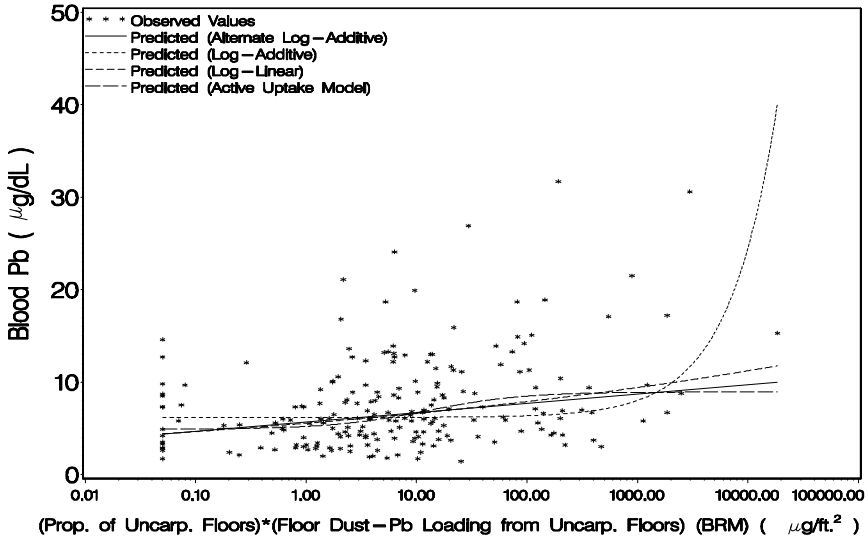


Figure G11-9. Bivariate Relationship Between Blood-Lead Concentration and Floor Dust-Lead Loading from Uncarpeted Surfaces (BRM Samples).

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.71 (0.0508)	0.08 (0.0164)	--	0.3435	0.0970	-180.35	205
Log-Additive	6.18 (0.2713)	0.00 (0.0009)	--	0.3651	0.0402	-186.61	205
Alternate Log-Additive	5.68 (0.2715)	0.44 (0.1007)	--	0.3467	0.0886	-181.30	205
Active Uptake	10.90 (2.4049)	1.58 (1.1784)	8.93 (0.9558)	0.3455	0.1008	-179.93	205



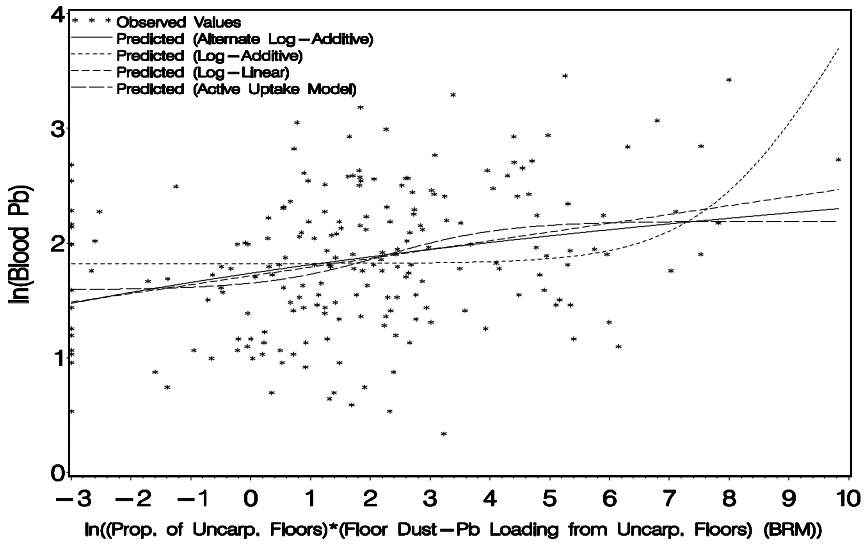


Figure G11-10. Bivariate Relationship Between Blood-Lead Concentration and (Proportion of Uncarpeted Floors Samples)*(Floor Dust-Lead Loading from Uncarpeted Floors)(BRM Samples).

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.20 (0.1530)	0.11 (0.0265)	--	0.3203	0.0923	-151.08	179
Log-Additive	6.06 (0.2848)	0.00 (0.0001)	--	0.3498	0.0085	-158.99	179
Alternate Log-Additive	2.46 (0.8656)	0.69 (0.1658)	--	0.3204	0.0920	-151.11	179
Active Uptake	9.11 (2.0871)	0.03 (0.0259)	9.34 (1.4739)	0.3197	0.1042	-149.91	179

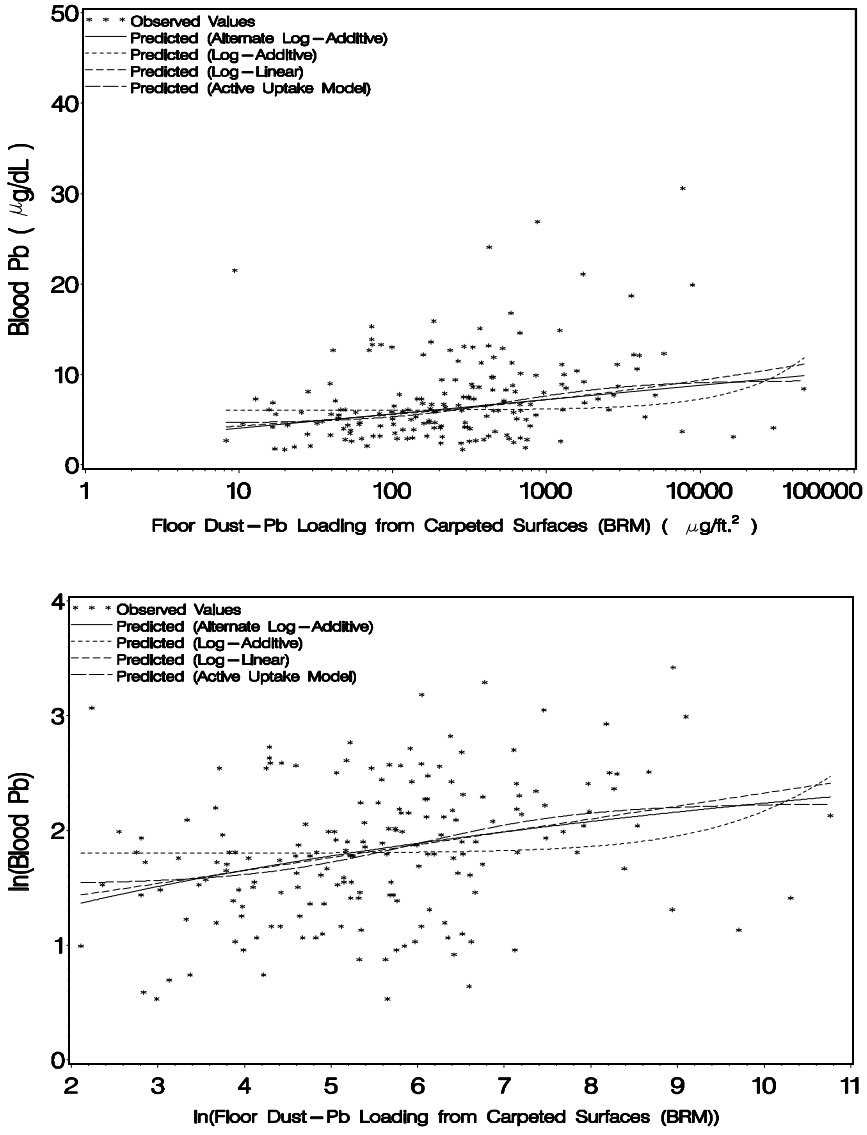


Figure G11-11. Bivariate Relationship Between Blood-Lead Concentration and Floor Dust-Lead Loading from Carpeted Surfaces (BRM Samples).

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.85 (0.0679)	0.00 (0.0137)	--	0.3804	0.0000	-190.81	205
Log-Additive	6.32 (0.2836)	0.00 (0.0001)	--	0.3795	0.0025	-190.56	205
Alternate Log-Additive	6.38 (0.4336)	0.00 (0.0876)	--	0.3804	0.0000	-190.81	205
Active Uptake	15.43 (6.6736)	0.02 (0.0398)	9.37 (2.5211)	0.3720	0.0317	-187.52	205

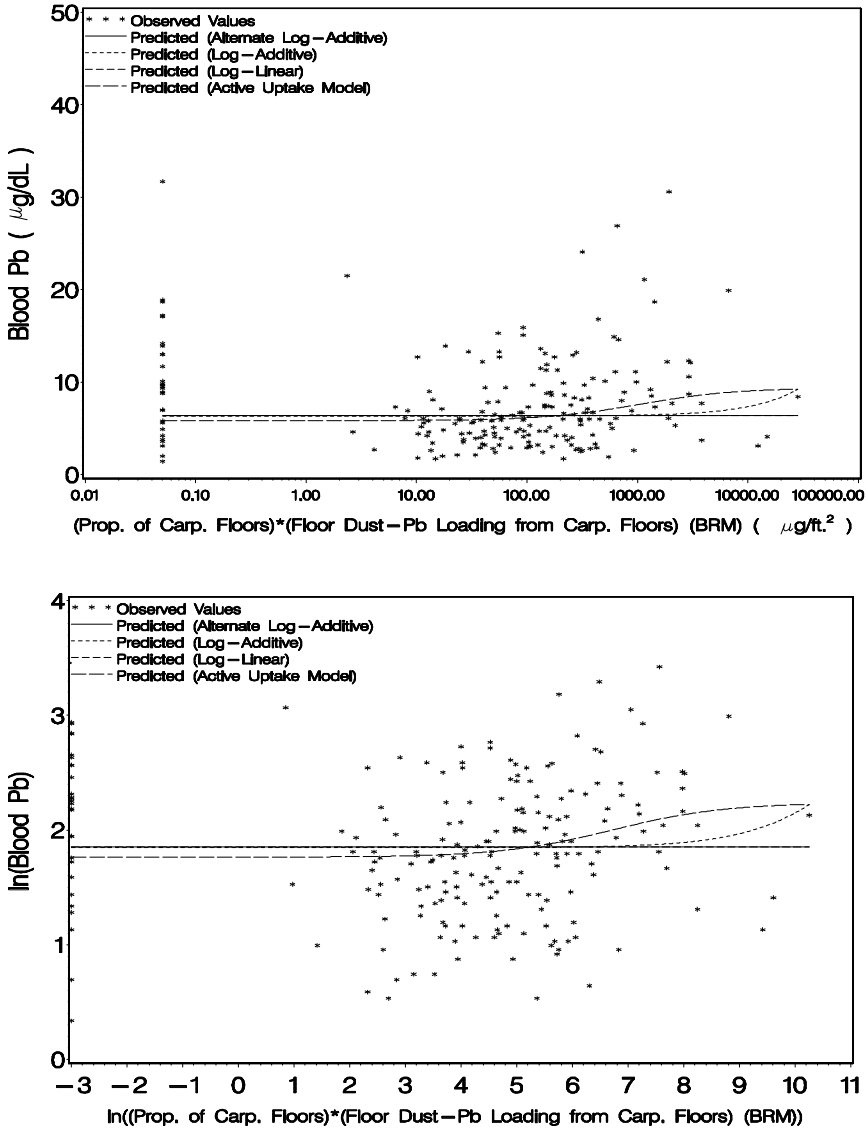


Figure G11-12. Bivariate Relationship Between Blood-Lead Concentration and (Proportion of Carpeted Floors Sampled)*(Floor Dust-Lead Loading from Carpeted Floors)(BRM Samples).

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.32 (0.1114)	0.09 (0.0177)	--	0.3382	0.1176	-171.75	197
Log-Additive	5.84 (0.2692)	0.00 (0.0000)	--	0.3460	0.0974	-173.98	197
Alternate Log-Additive	3.46 (0.6138)	0.51 (0.1123)	--	0.3428	0.1055	-173.09	197
Active Uptake	10.15 (1.7822)	0.00 (0.0041)	11.29 (2.0248)	0.3389	0.1249	-170.94	197

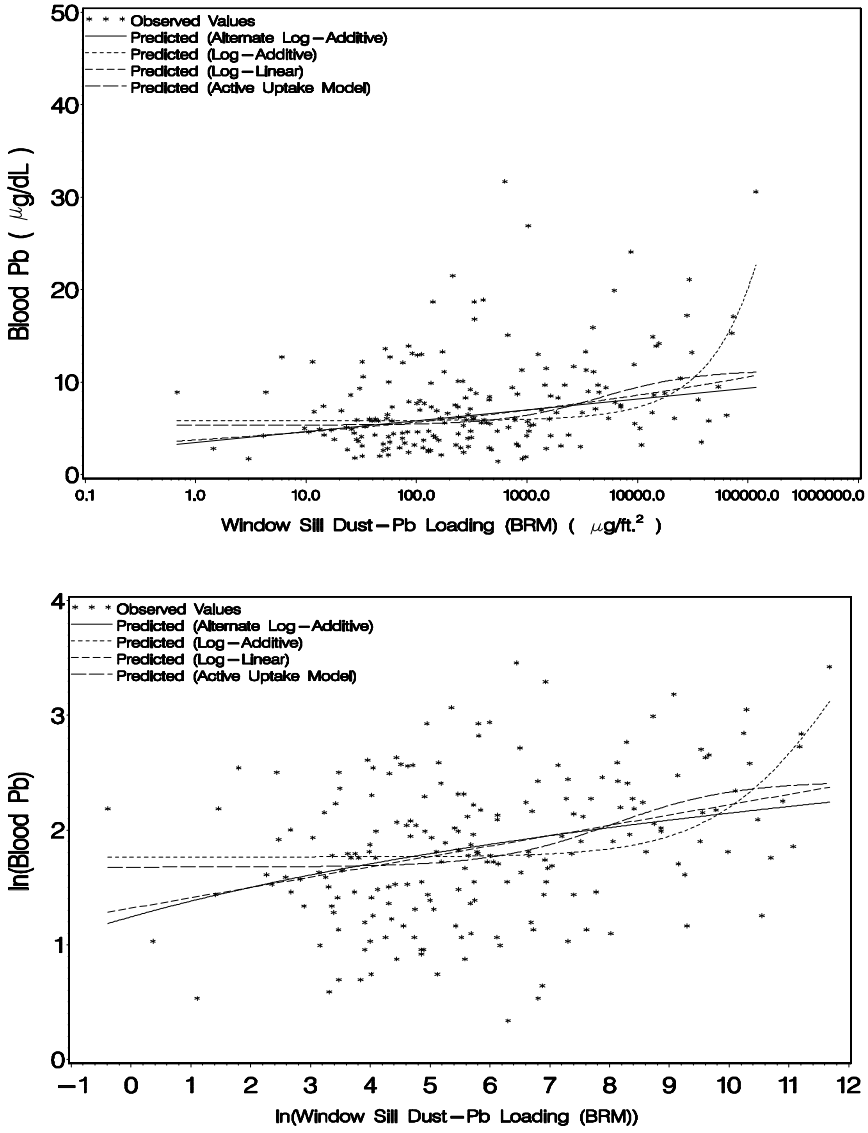


Figure G11-13. Bivariate Relationship Between Blood-Lead Concentration and Window Sill Dust-Lead Loading (BRM Samples).

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.08 (0.1444)	0.08 (0.0138)	--	0.3366	0.1402	-164.29	189
Log-Additive	5.54 (0.2938)	0.00 (0.0000)	--	0.3530	0.0983	-168.78	189
Alternate Log-Additive	2.41 (0.6968)	0.40 (0.0758)	--	0.3414	0.1279	-165.63	189
Active Uptake	9.35 (1.6300)	0.00 (0.0001)	9.83 (1.4255)	0.3331	0.1581	-162.30	189

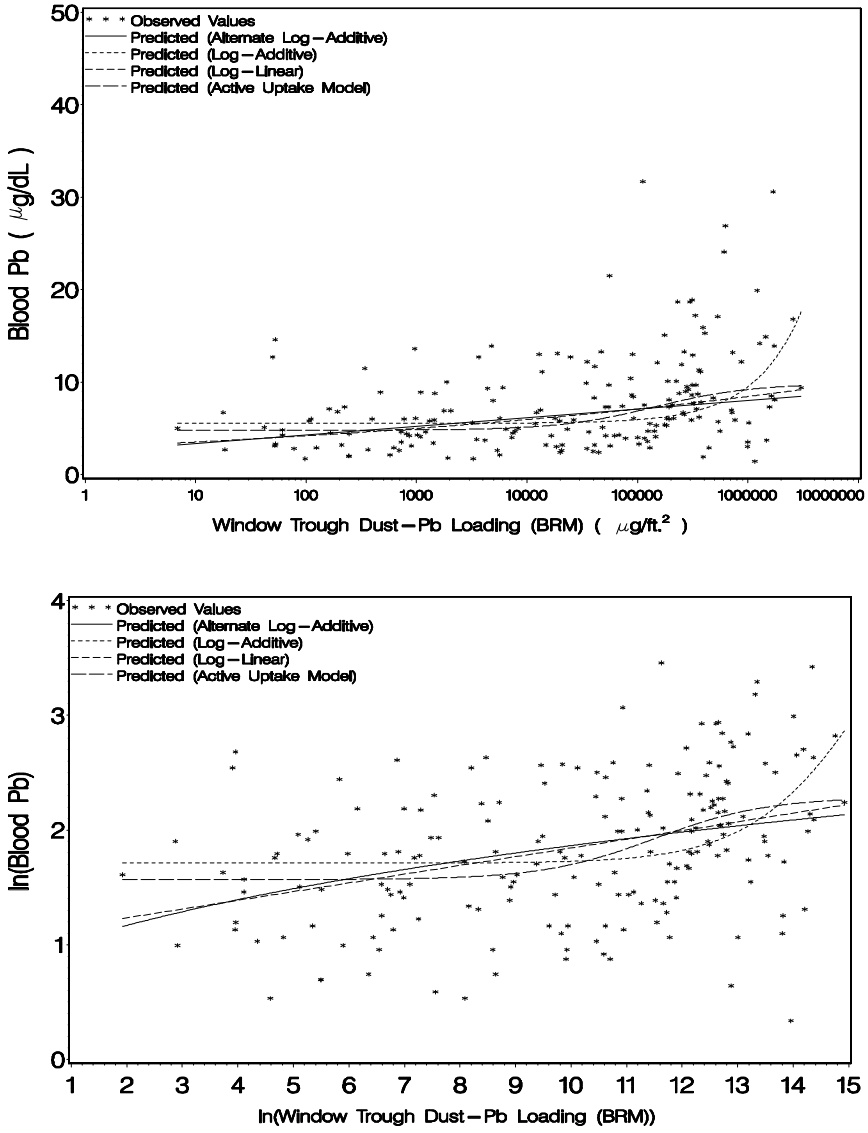


Figure G11-14. Bivariate Relationship Between Blood-Lead Concentration and Window Well Dust-Lead Loading (BRM Samples).

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	0.73 (0.2216)	0.17 (0.0330)	--	0.3410	0.1288	-162.86	186
Log-Additive	5.31 (0.3700)	0.00 (0.0003)	--	0.3647	0.0683	-169.11	186
Alternate Log-Additive	0.16 (1.0046)	0.97 (0.1655)	--	0.3414	0.1278	-162.97	186
Active Uptake	6.53 (1.4373)	0.01 (0.0092)	10.40 (2.1910)	0.3443	0.1299	-162.75	186

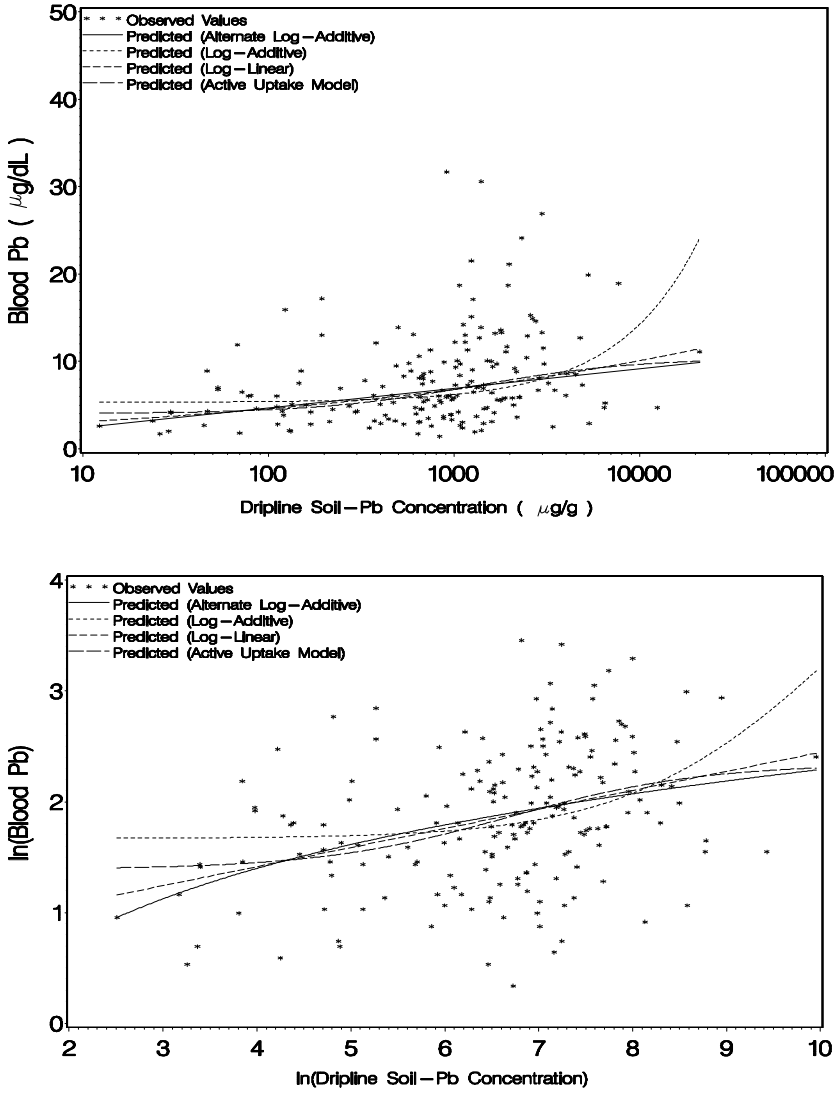


Figure G11-15. Bivariate Relationship Between Blood-Lead Concentration and Dripline Soil-Lead Concentration.

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	0.94 (0.2941)	0.16 (0.0518)	--	0.2419	0.1044	-60.71	87
Log-Additive	5.65 (0.4364)	0.00 (0.0008)	--	0.2528	0.0642	-62.62	87
Alternate Log-Additive	0.74 (1.7617)	1.03 (0.3294)	--	0.2422	0.1032	-60.77	87
Active Uptake	8.23 (2.5304)	0.02 (0.0293)	11.15 (4.5624)	0.2475	0.1050	-60.69	87

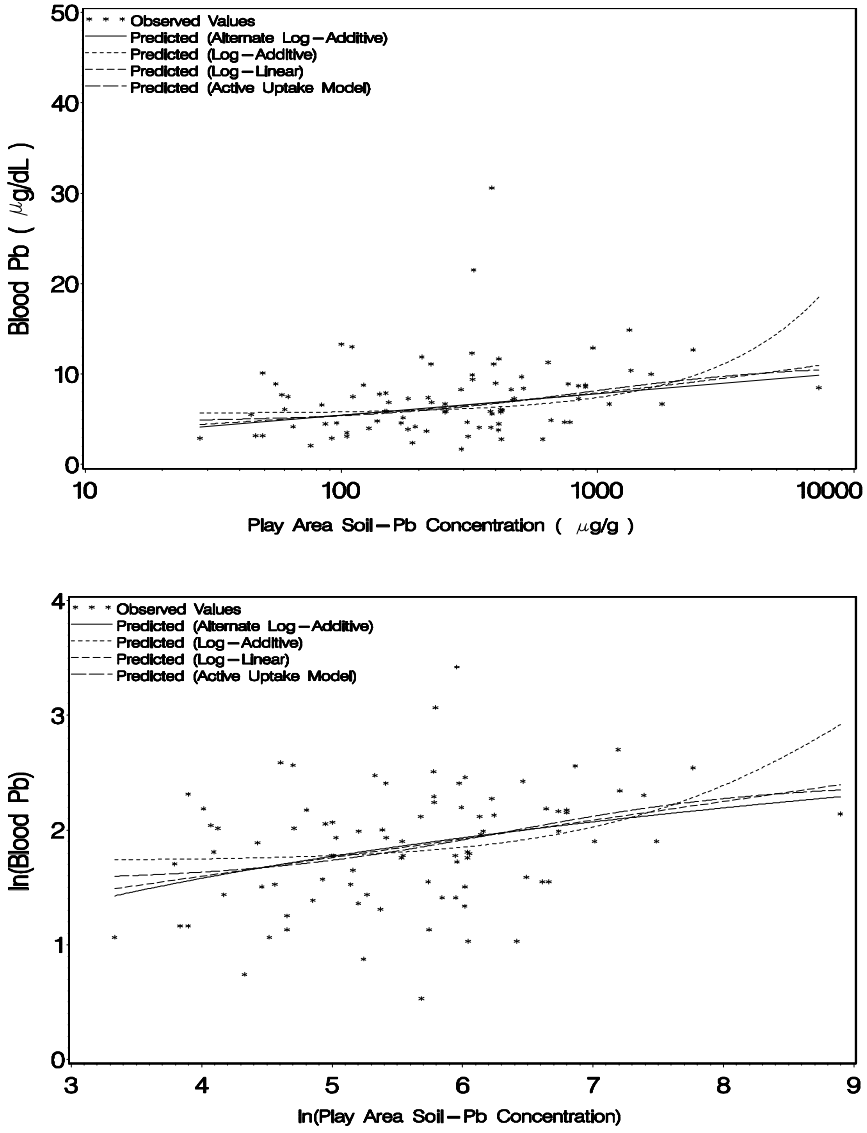


Figure G11-16. Bivariate Relationship Between Blood-Lead Concentration and Play Area Soil-Lead Concentration.

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	3.15 (0.7032)	0.66 (0.1487)	--	0.3483	0.0903	-180.00	205
Log-Additive	6.14 (0.2867)	0.00 (0.0002)	--	0.3752	0.0203	-187.53	205
Alternate Log-Additive	1.31 (0.1269)	0.11 (0.0238)	--	0.3476	0.0922	-179.79	205
Active Uptake	9.70 (2.1860)	0.06 (0.0491)	9.29 (1.2981)	0.3488	0.0981	-179.14	205

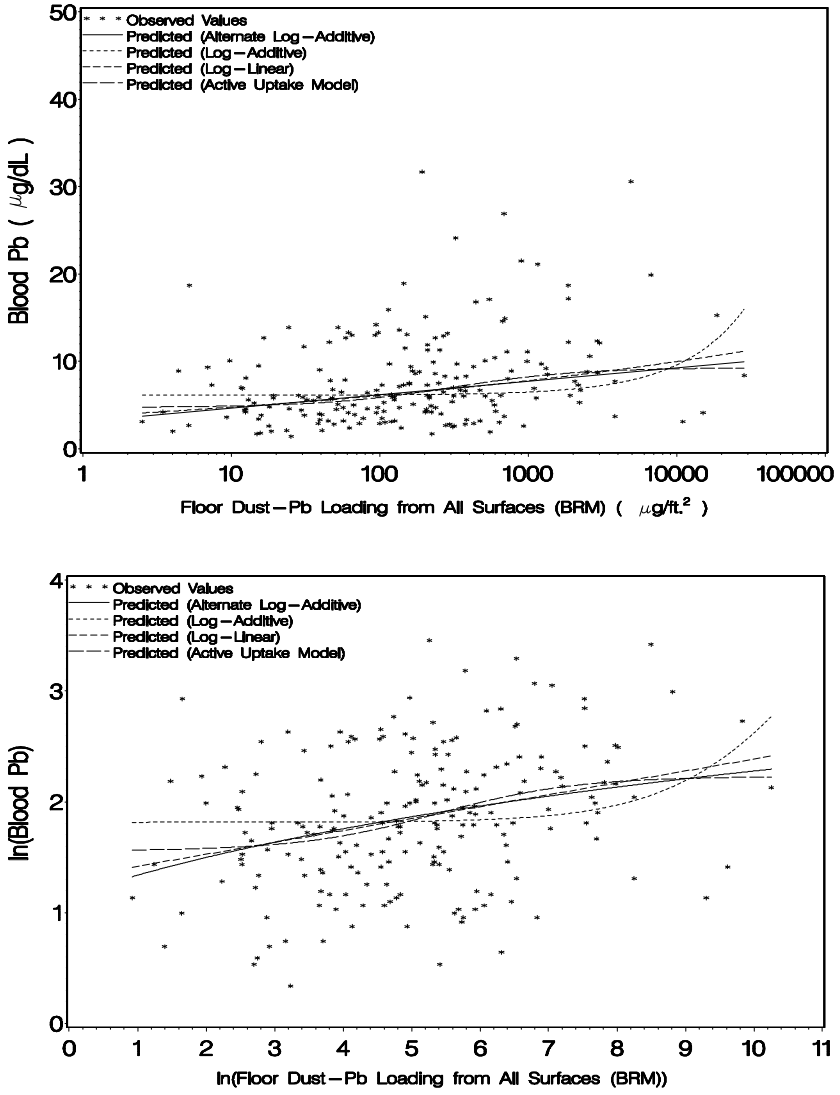
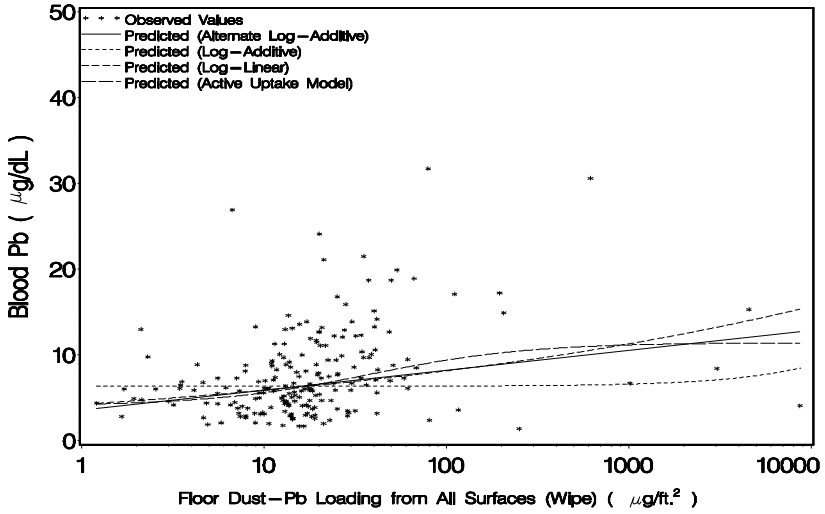


Figure G11-17. Bivariate Relationship Between Blood-Lead Concentration and the Total Effect of Floor Dust-Lead Loading from All Surfaces (Carpeted or Uncarpeted) (BRM Samples).

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	3.57 (0.6925)	1.01 (0.2547)	--	0.3511	0.0771	-182.59	203
Log-Additive	6.36 (0.2776)	0.00 (0.0005)	--	0.3800	0.0011	-190.70	203
Alternate Log-Additive	1.45 (0.1109)	0.14 (0.0358)	--	0.3531	0.0717	-183.18	203
Active Uptake	6.33 (1.6813)	0.47 (0.2797)	11.44 (2.3671)	0.3438	0.1051	-179.44	203



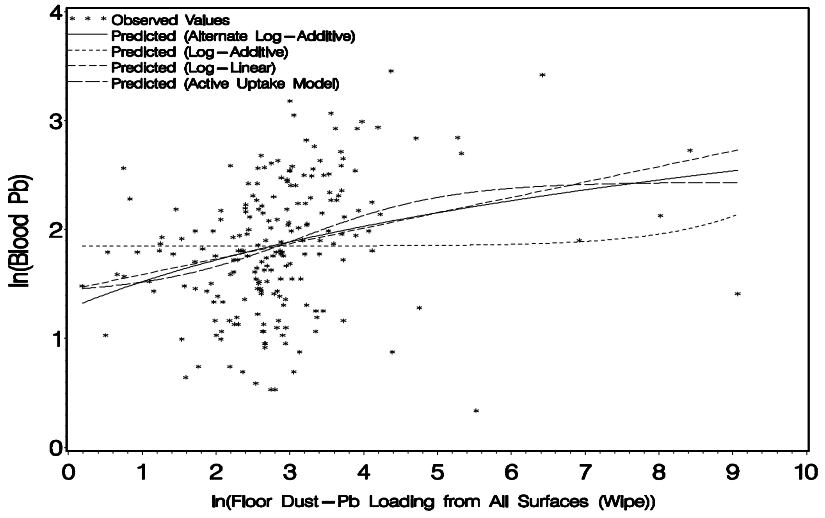


Figure G11-18. Bivariate Relationship Between Blood-Lead Concentration and the Total Effect of Floor Dust-Lead Loading from All Surfaces (Carpeted or Uncarpeted) (Wipe Samples).

Statistical Model	Parameter Estimates and (Associated Standard Errors)			^{2}	R^2	Estimated Log-Likelihood	Number of Observations
	β_0 s.e. (β_0)	β_1 s.e. (β_1)	θ_{HSP} s.e. (θ_{HSP})				
Log-Linear	1.82 (0.0437)	0.33 (0.1058)		0.3628	0.0463	-185.95	205
Log-Additive	6.15 (0.2696)	2.67 (1.0765)		0.3630	0.0457	-186.02	205
Alternate Log-Additive	6.15 (0.2696)	2.67 (1.0765)		0.3630	0.0457	-186.02	205
Active Uptake	6.15 (11.3986)	2.66 (13.2990)	2.49E8 (1.8727E16)	0.3666	0.0457	-186.02	205

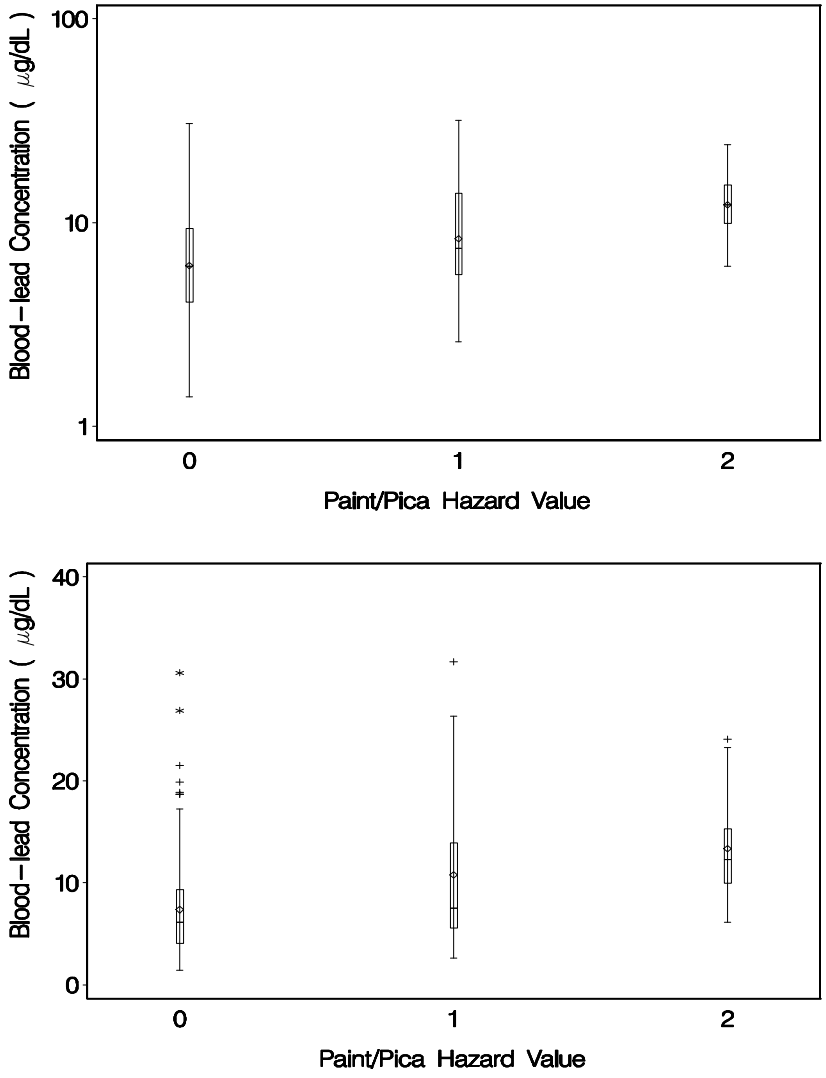


Figure G11-19. Bivariate Relationship Between Blood-Lead Concentration and Paint/Pica Hazard Variable (Interior).

G12: Appendix on Regression Diagnostics

Regression Diagnostics

This section of the appendix describes the diagnostic analyses performed as part of development of a multimedia exposure model using data from the Rochester Lead-In-Dust Study. Through the use of regression diagnostics, the adequacy of fit of the various candidate models developed (including the multi-media predictive model) to the data observed can be determined, and model assumptions can be verified. Results are presented for the final chosen model in particular, for which the following regression diagnostic "stages" were performed:

1. A normal quantile plot of the residuals was created. A normal quantile plot which can best be described by a straight line indicates that residuals (errors) are approximately normally distributed, as assumed. The quantile plot given in Figure G12-1 can best be described by a straight line, and therefore the assumption of normal errors is satisfied.

2. Residual values were plotted versus predicted values. This scatterplot could indicate signs of nonconstant variance (if points spread out or tighten up as you move from left to right) or nonlinearity (if points look quadratic or bow-shaped). A scatterplot exhibiting no pattern indicates no such problems. Similarly, plots of residuals versus predictors should indicate no discernible pattern. A plot of residuals versus predicted values is given in Figure G12-2. A plot of residuals versus predictor variables are given in Figure G12-3. Note that none of these plots indicate any relationship and each resembles a somewhat random scattering of points.

3. A plot of Cook's distance and DFFITS (both measures of influence) versus studentized residuals (a measure of how far an observation deviates from the modeled relationship) can indicate potential outliers - points with undue influence and points lying far outside the model's prediction. A plot of these two influence statistics are given in Figure G12-4. Each of these plots point to two possible outliers: observations with Child Identification Number (CID) 00166 and 04072. The observation with CID 00166 is also the observation with the lowest PbB level, while the observation with CID 04072 has the largest PbF level and the fifth smallest PbS level, and thus may require further examination. Note that DFFITS and Cook's distance are related to the studentized residuals and by definition are themselves similar, so observable patterns in these plots indicate nothing. However, typically those points with large studentized residuals (larger than 3 in absolute value) or DFFITS (larger than 1 in absolute value), or Cook's distance (larger than 1) possibly require further examination.

4. For a closer examination of how points influence model parameter estimates, the models were fit while excluding a single point at a time. Analysis of the coefficients adjusted for their standard error (intercept, and coefficients of PbS, PbF, PbW and PbP), including plots, can provide information about the influence of specific observations. Plots of the scaled measure of change in each parameter estimate are provided in the scatterplot matrix of Figure G12-5. Typically, values exceeding 1 in absolute value are suspect points. Note that none of the points in the multi-media

predictive model analysis is suspect by this criteria. Table G12-1 below provides the parameter estimates while excluding the potential outliers flagged in stage (3).

Table G12-1. Influence of Possible Outlying Observations

Param.	Description	Estimate (deleting CIDs 00166, 04072)	Estimate (deleting CID 00166)	Estimate (Model with no deletions)
β_0	Intercept	0.427484 (0.234447)	0.403628 (0.234713)	0.417648 (0.240347)
β_1	log (PbS): Drip-line Soil-Lead Concentration (fine soil fraction)	0.101146 (0.035592)	0.115042 (0.034462)	0.114038 (0.035294)
β_2	PbP: Indicator of Interior Paint/Pica Hazard	0.229457 (0.097897)	0.236655 (0.098118)	0.248043 (0.100421)
β_3	log (PbF): Area-Weighted Arithmetic Mean (Wipe) Dust-Lead Loading from Any Floor (Carpeted or Uncarpeted)	0.119694 (0.044423)	0.090483 (0.039976)	0.066338 (0.040151)
β_4	log (PbW): Area-weighted Arithmetic Mean (Wipe) Dust-Lead Loading from Window Sills	0.075433 (0.035178)	0.077318 (0.035277)	0.087010 (0.035987)
R^2	Coefficient of Determination	23.98%	23.23%	21.67%
	Root Mean-Square Error (Residual Error)	0.54670	0.54861	0.56188

This table indicates that excluding these points changes the parameter estimates only slightly.

5. Partial regression leverage plots were created for the environmental measures of lead exposure: dripline soil, floor dust from carpeted and uncarpeted floors, paint/pica hazard, and window sill dust. A partial regression leverage plot that exhibits a linear relationship between blood-lead and the variable under consideration is indicative of a linear relationship between blood lead and the environmental measure of lead exposure while controlling for all the other variables in the model. The partial regression leverage plots given in Figure G12-6 indicate adjusted linear relationships for the lead-exposure variables included in the log-linear multimedia exposure model fitted to the data from the Rochester Lead-in-Dust Study. Note that a partial regression leverage plot is produced by plotting the residuals from a regression of the response variable $(LPbB_{ijk})$ on all predictor variables excluding the lead exposure variable under consideration, versus the residuals from a regression of the lead exposure variable under consideration on the remaining predictor variables.

6. Partial R^2 comparisons can be made between predictor variables included in the model. A high partial R^2 indicates greater importance in predicting blood-lead concentration. Table G12-2 below provides the coefficient of determination (R^2) for a series of models in which one of the four predictor variables is excluded from the log-linear model. The additional amount of variability in blood-lead concentrations explained by the excluded predictor variable once added to the model is also provided.

Table G12-2. Partial R-squared Comparisons.

Variable Excluded from the Model	Coefficient of Determination(R^2)	Partial Coefficient of Determination (Partial R^2)[a]	Additional Variability Explained = (21.67% - R^2)[b]
Paint/Pica Hazard	18.93%	3.38%	2.74%
Floor Dust-Lead	20.44%	1.54%	1.23%
Dripline Soil-Lead	16.97%	5.67%	2.63%
Window Sill Dust-Lead	19.04%	3.25%	4.70%

[a] Partial R^2 gives the contribution to the percent variation explained by adding in the variable of interest. It is calculated as: $\frac{R^2\ (FULL) - R^2\ (REDUCED)}{1 - R^2\ (REDUCED)}$.

[b] 21.67% denotes the coefficient of determination (R^2) for the full multi-media predictive model.

The multi-media predictive model explains 21.67% of the variability in childhood blood-lead concentrations. Exposure from soil is the best predictor of blood-lead concentration, with the highest partial R^2 of around five percent.

7. An analysis into the effects of collinearity using several methods was conducted during the development of the multi-media predictive model. Issues pertaining to collinearity and strong correlation among potential lead-exposure predictor variables had a prominent role in the variable selection for the multi-media predictive model. Estimates of the tolerance statistic and variance inflation factor associated with each predictor variable in the model are provided in Table G12-3, together with a single value decomposition for the design matrix of observed predictor variables in the Rochester Study.

   To aid in the interpretation of these collinearity diagnostics, note that a large condition index indicates the data are ill-conditioned, or when extremely large, that parameter estimates are subject to substantial numerical error. A collinearity problem occurs whenever a variable with a high condition index is also a chief contributor to the variability between two or more variables.

   Variance inflation factors measure how much the variability associated with a particular parameter estimate is inflated due to collinearity between the predictors in a regression model. Although no formal criteria exists for establishing a critical variance inflation factor, it is common practice to associate a condition index of 10 with the notion that weak dependencies may be starting to affect the regression estimates. Condition indices of 30 to 100 indicate moderate to strong dependencies, and indices of greater than 100 indicate serious collinearity problems. The number of condition indices in the critical range indicates the number of near dependencies contributing to the collinearity problem.

Finally, another collinearity diagnostic is the condition number , defined by = (largest eigenvalue / smallest eigenvalue)^{½}, where large values suggest collinearity.

Table G12-3. Collinearity Diagnostics

Index	Eigenvalue	Condition Index	PbF	PbW	PbS	PbP
			Proportion of Variability Explained			
1	1.70803	1.00000	0.1395	0.1659	0.1295	0.0380
2	0.95248	1.33912	0.0264	0.0360	0.0013	0.9450
3	0.81482	1.44783	0.4116	0.0009	0.6107	0.0105
4	0.52466	1.80430	0.4225	0.7972	0.2585	0.0065
			Tolerance			
			0.820436	0.736608	0.855715	0.976984
			Variance Inflation			
			1.218864	1.357574	1.168613	1.0235585

Note that the largest condition index in Table G12-3 is 1.8, and the largest inflation factor is 1.36 (PbW). Therefore, the multi-media predictive model (in its current form) does not appear to suffer from a severe collinearity problem, nor does it appear to be ill-conditioned (numerically unstable or fragile). The following matrix contains the correlation coefficients among the four predictor variables used in the multi-media predictive model. The coefficients are based on a sample size of 179 children/ households included in the current model.

	PbF	PbW	PbS	PbP
PbF	1.000	0.417	0.186	0.110
PbW	0.417	1.000	0.370	0.101
PbS	0.186	0.370	1.000	0.119
PbP	0.110	0.101	0.119	1.000

Plots are provided in Figure G12-7 of each continuous predictor variable versus another continuous predictor variable, where each observation is coded for values of the paint/pica hazard variable (0, 1 or 2). These plots provide insight into the range of possible values over which the multi-media predictive model was constructed, and over which inferences can be drawn.

Based on the regression diagnostics on the multi-media predictive model it was concluded that:

- no influential or outlying points should be deleted from the analysis,
- the model developed fits the data observed,
- model assumptions are verified, and
- the model does not appear to suffer from a severe problem with collinearity.

Table G12-4. Formulae of Regression Diagnostic Statistics

Variable and Description	Formula
Predicted	$\hat{y}_i = \beta_0 + \beta_1 x_{1,i} + \ldots + \beta_k x_{k,i}$
Residual	$\hat{e}_i = \hat{y}_i - y_i$
Leverage (or hat matrix diagonals)	$h_i = x_i (X'X)^{-1} x_i'$
Quantity x calculated without jth observation	$x_{(j)}$
Externally Studentized Residual	$r_i = \left(\frac{\hat{e}_i}{\hat{}_{(i)} \sqrt{1 - h_i}} \right)$
Cook's Distance	$D_i = \frac{\hat{e}_{(i)}}{\hat{} \sqrt{h_i}}$
DFFITS	$DFFITS_i = \frac{\hat{e}_{(i)}}{\hat{}_{(i)} \sqrt{h_i}}$
COVRATIO	$C_i = \frac{\det(\hat{}^2_{(i)} (X'_{(i)} X_{(i)})^{-1})}{\det(\hat{}^2 (X'X)^{-1})}$
DFBETA	$DFBETA_{i,j} = \frac{\beta_j - \beta_{j(i)}}{\hat{}_i \sqrt{(X'X)^{-1}_{jj}}}$

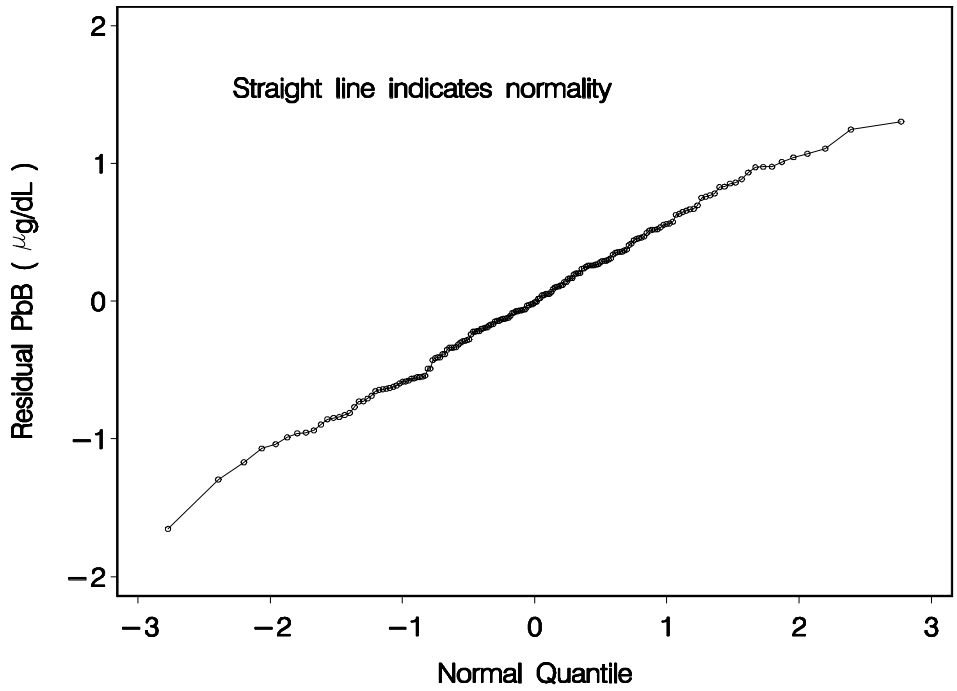


Figure G12-1. Quantile Plot of Residuals.

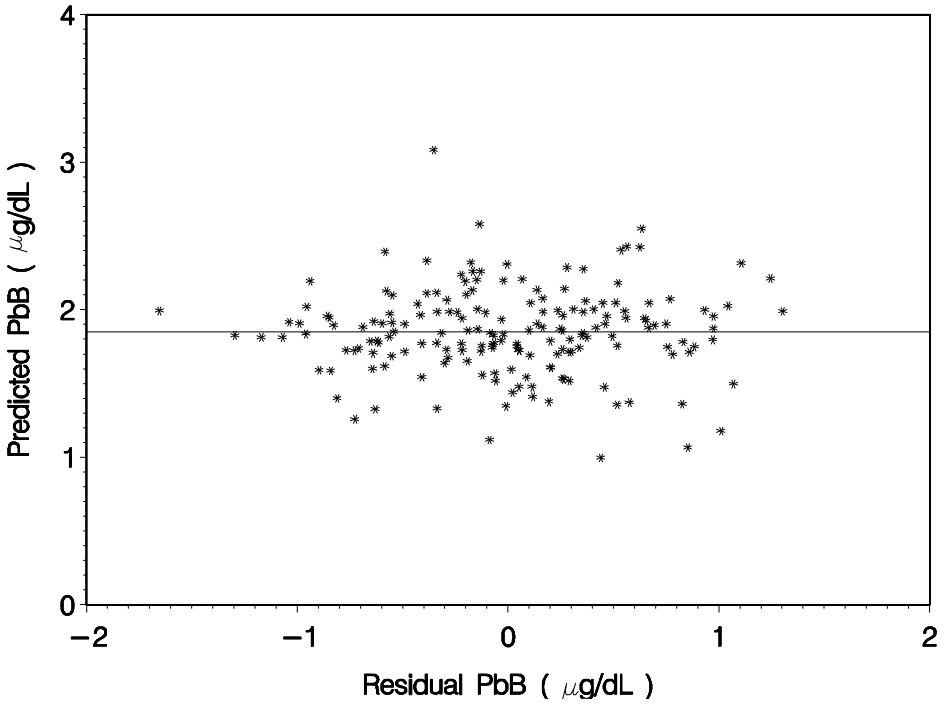


Figure G12-2. Plot of Residuals versus Predicted Values.

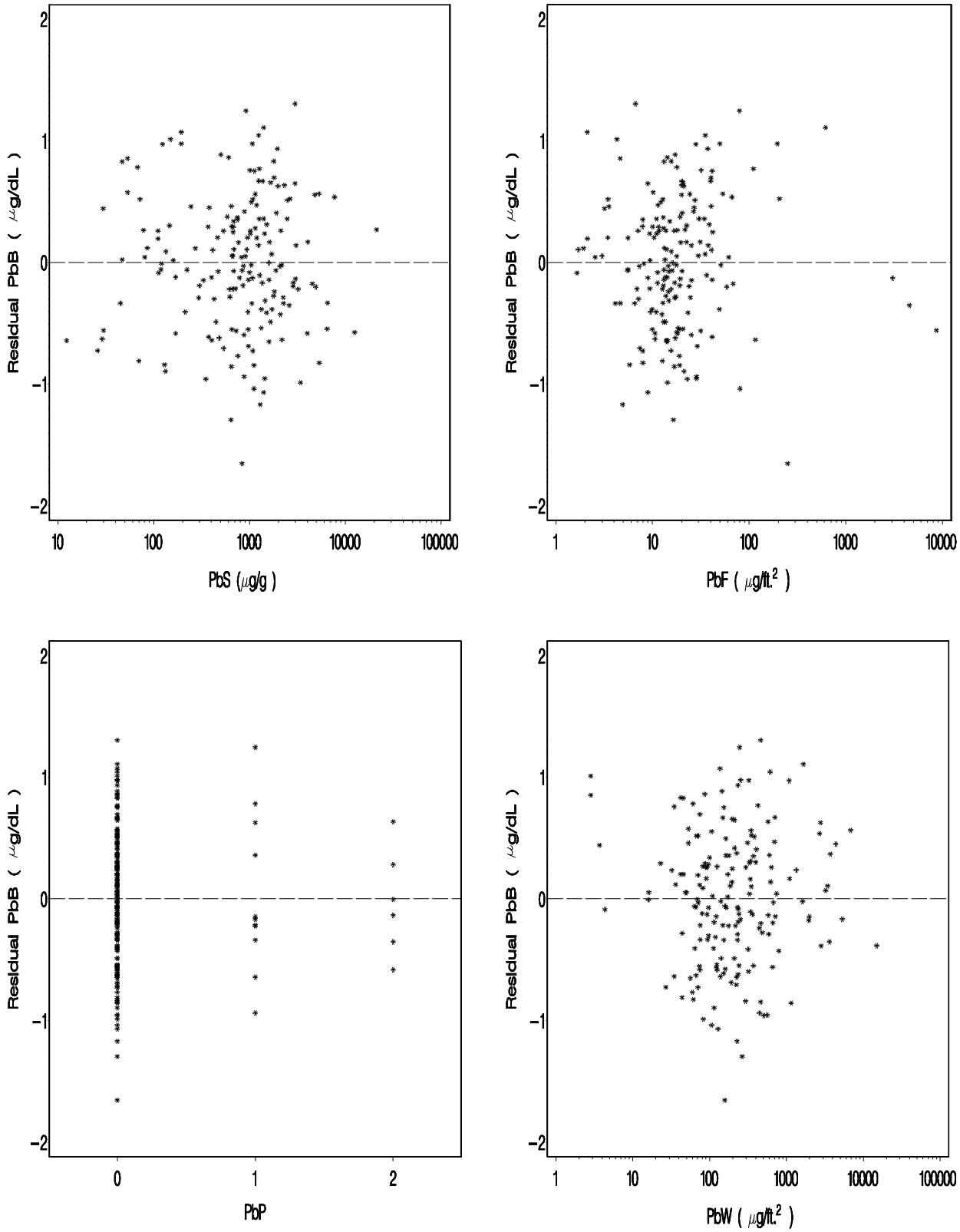


Figure G12-3. Plots of Residuals versus Predictors.

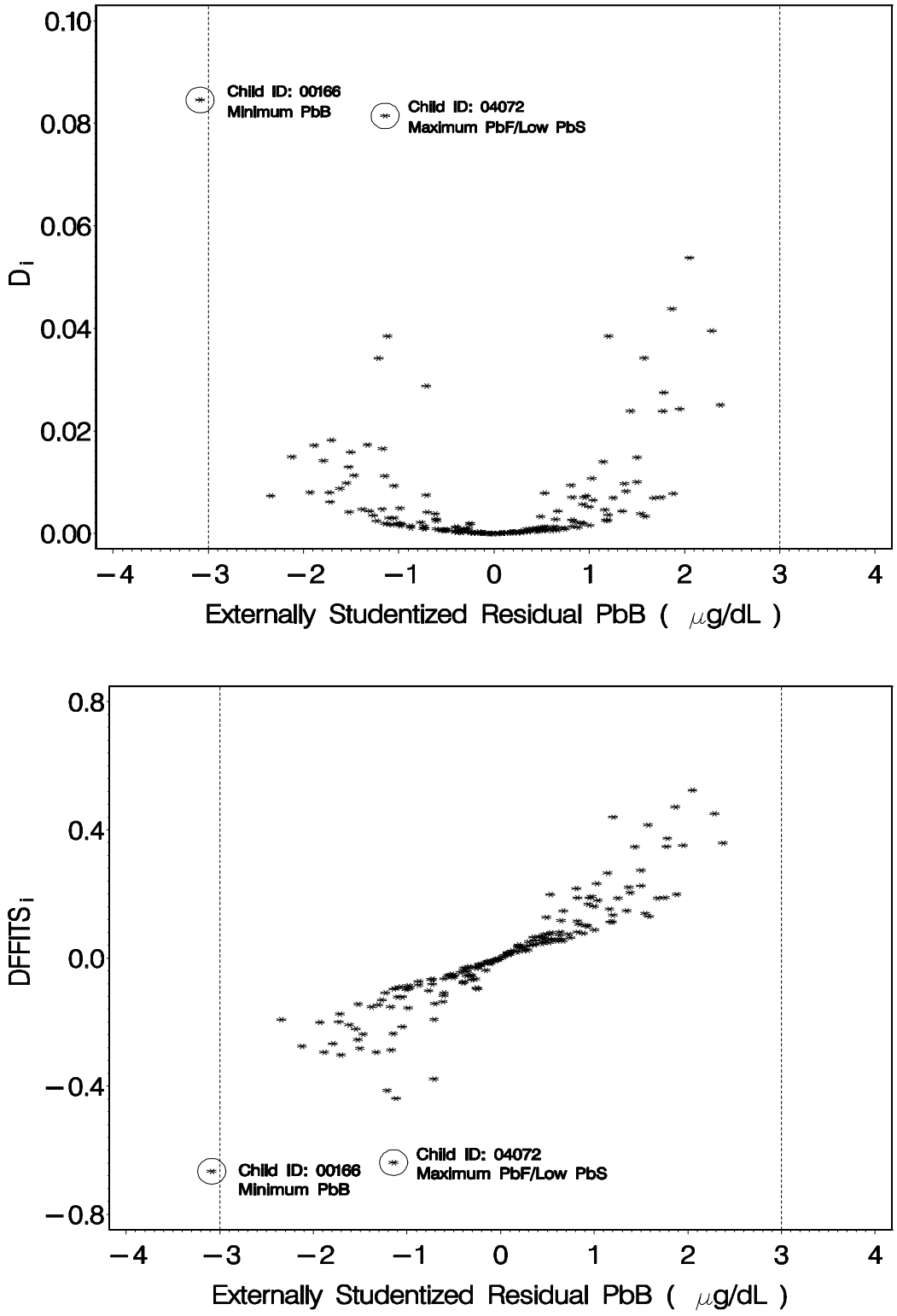


Figure G12-4. Plots of Influence Statistics (Cook's Distance (D_i) and $DFFITS_i$) versus Externally Studentized Residuals.

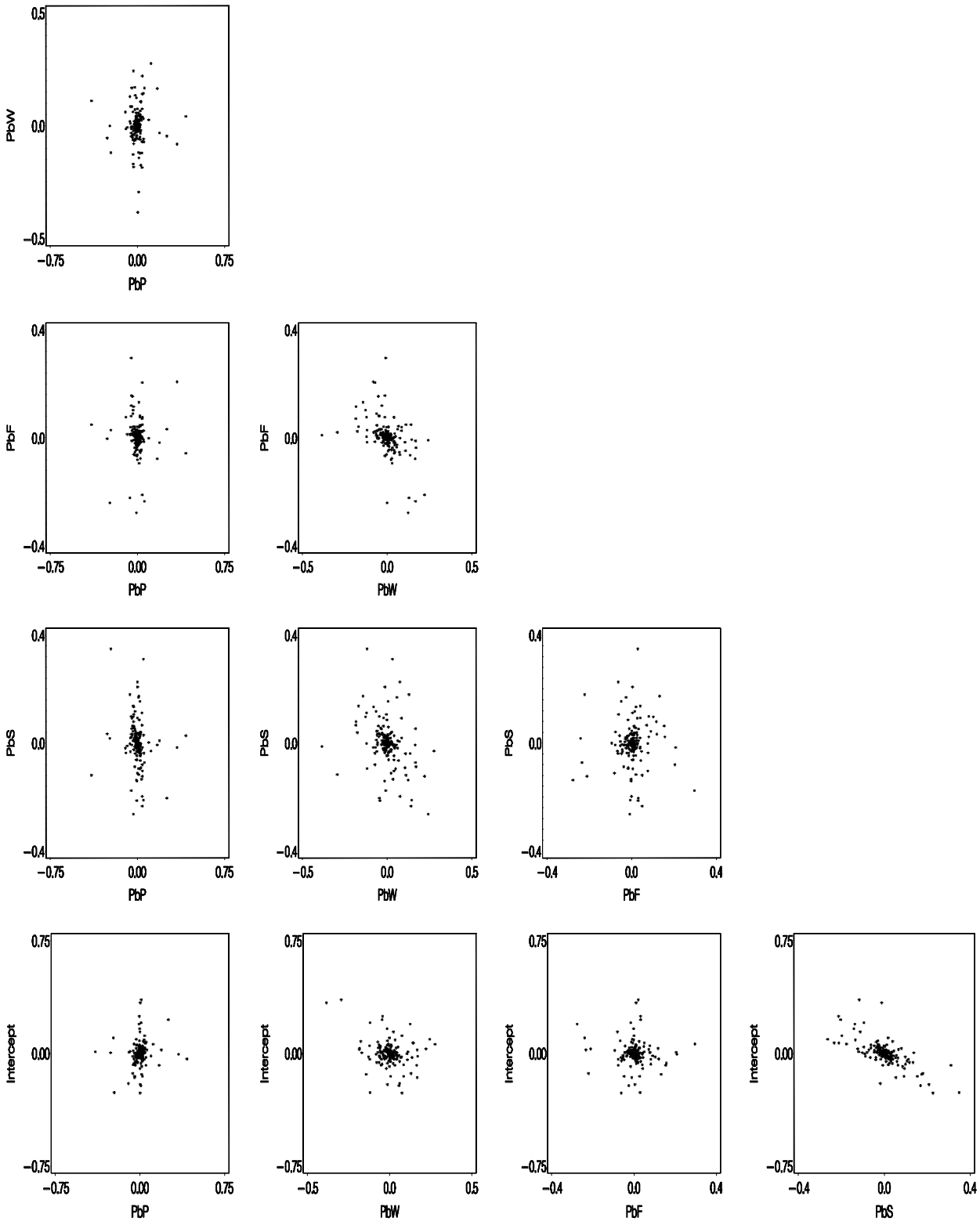


Figure G12-5. Plot of Changes in Parameter Estimates Relative to Standard Error for Intercept and Coefficients of PbS, PbW, PbF and PbP.

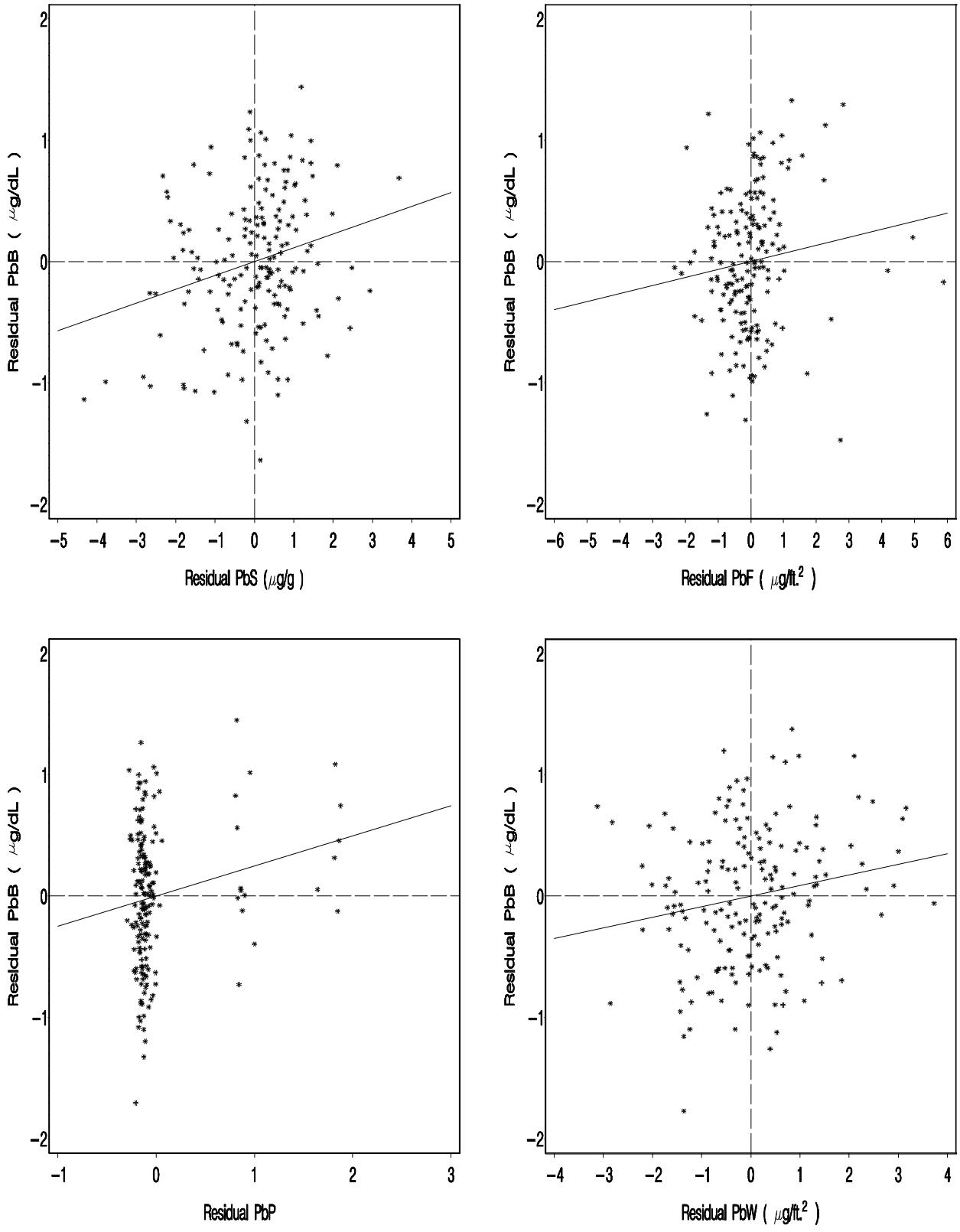


Figure G12-6. Partial Leverage Regression Plots.

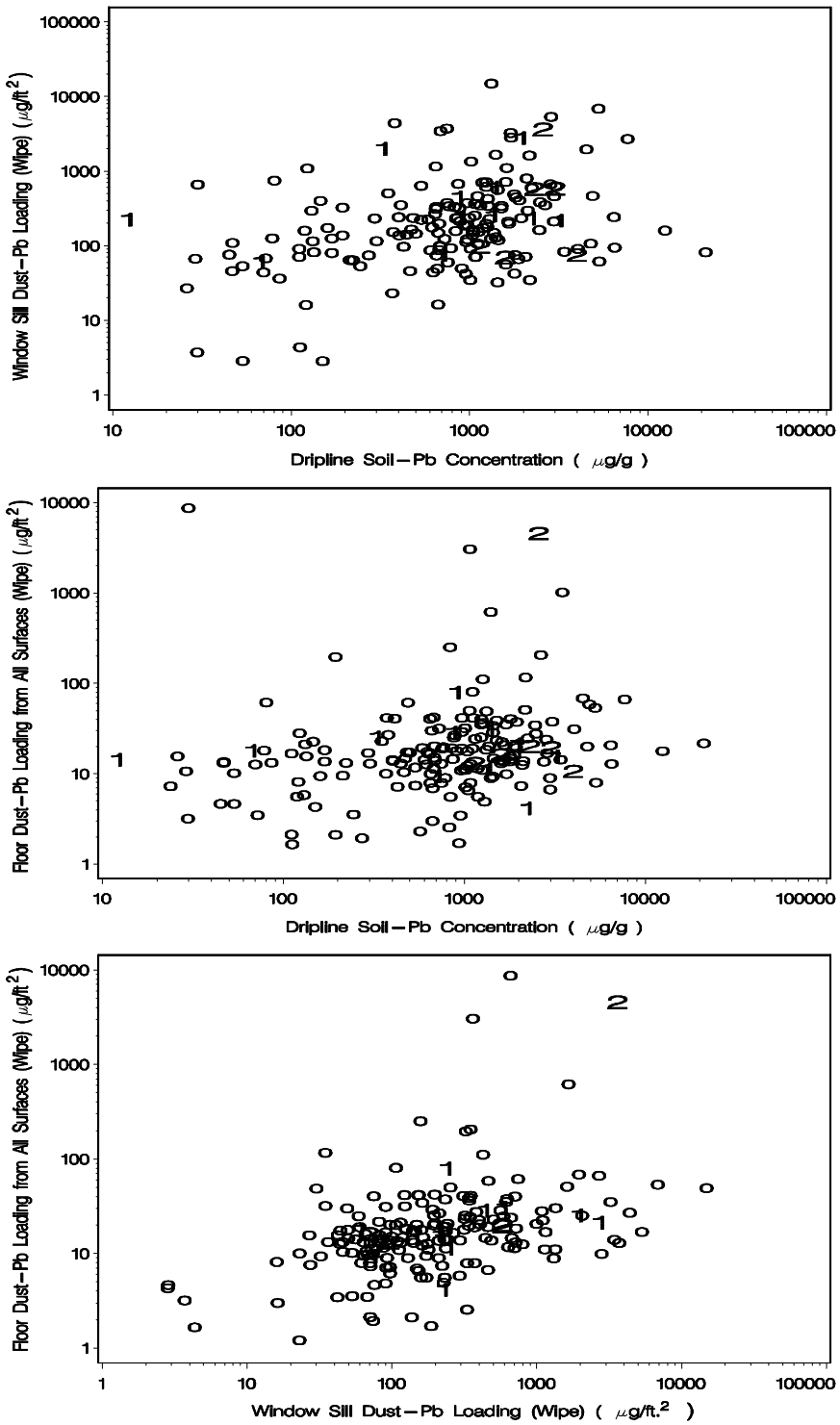


Figure G12-7. Plots of One Predictor Variable versus Another Predictor Variable Coded for Values of Paint/Pica Hazard Variable.

G13: Appendix on Parameter Estimates for Candidate Multimedia Exposure Models

Parameter Estimates for Candidate Multimedia Exposure Models

Table G13-1. Comparison of Parameter Estimates (Standard Errors) Obtained from Five Competing Statistical Models Which Regress Blood-Lead Concentration on Measures of Drip-Line Soil-Lead Concentration, Floor Dust-Lead Loading, Paint/Pica Hazard and Race.

Parameter	Parameter Estimates(Standard Errors) for the Five Statistical Models				
	Log-Linear[a]	Log-Additive	Alternative[a] Log-Additive	Active Uptake	Active/ Passive Uptake
β_0 Intercept	0.533 (0.200)	4.279 (0.312)	-0.593 (0.924)	4.610 (0.768)	4.610 (2.940)
β_1 Floor	0.070 (0.032)	0.0002 (0.0003)	0.249 (0.186)	0.0002 (0.0005)	0.0002 (0.0006)
β_2 Soil	0.134 (0.029)	0.0006 (0.0002)	0.772 (0.148)	0.0011 (0.0008)	0.0011 (0.0019)
β_3 Paint/Pica	0.219 (0.090)	1.798 (0.897)	1.667 (0.819)	2.570 (2.491)	2.571 (5.329)
β_4 Black	0.524 (0.076)	3.642 (0.605)	3.422 (0.569)	5.409 (3.832)	5.410 (10.451)
θ[b]				36.086 (52.295)	36.075 (415.797)
$F_{Passive}$[c]					0.000 (5.550)
^{2}	0.256	0.286	0.262	0.289	0.289
R^2	0.357	0.280	0.341	0.281	0.281
ln( )[d]	-134.595	-145.104	-136.920	-145.011	-145.029

[a] In the implementation of the log-linear and alternate log-additive models, the categorical variables Paint/Pica and Black were not log-transformed.

[b] The parameter θ appears in the Active Uptake and Active/Passive Uptake Models described in Section G3.0.

[c] The parameter $F_{Passive}$ appears in the Active/Passive Uptake Model described in Section G3.0.

[d] Ln( ) represents the log-likelihood of the observed Rochester data given each model, and can be used to assess the plausibility of one model in comparison to another, as described in Section G4.3.

Table G13-2. Comparison of Parameter Estimates (Standard Errors) Obtained from Five Competing Statistical Models Which Regress Blood-Lead Concentration on Measures of Drip-Line Soil-Lead Concentration, Floor Dust-Lead Loading, and Paint/Pica Hazard.

Parameter	Parameter Estimates(Standard Errors) for the Five Statistical Models				
	Log-Linear[a]	Log-Additive	Alternative[a] Log-Additive	Active Uptake	Active/ Passive Uptake
β_0 Intercept	0.608 (0.224)	5.221 (0.361)	-0.494 (1.079)	6.424 (1.323)	6.424 (1.763)
β_1 Floor	0.089 (0.036)	0.00001 (0.0004)	0.474 (0.242)	0.00006 (0.0008)	0.00006 (0.0008)
β_2 Soil	0.146 (0.033)	0.0008 (0.0002)	0.834 (0.177)	0.009 (0.006)	0.009 (0.010)
β_3 Paint/Pica	0.252 (0.101)	2.434 (1.075)	2.131 (0.985)	21.857 (34.860)	21.857 (43.079)
θ[b]				11.315 (2.760)	11.315 (6.623)
$F_{Passive}$[c]					0.000 (0.054)
^{2}	0.321	0.354	0.324	0.340	0.340
R^2	0.189	0.105	0.182	0.149	0.149
ln( )[d]	-156.216	-165.340	-156.979	-160.687	-160.702

[a] In the implementation of the log-linear and alternate log-additive models, the categorical variable Paint/Pica was not log-transformed.

[b] The parameter θ appears in the Active Uptake and Active/Passive Uptake Models described in Section G3.0.

[c] The parameter $F_{Passive}$ appears in the Active/Passive Uptake Model described in Section G3.0.

[d] Ln( ) represents the log-likelihood of the observed Rochester data given each model, and can be used to assess the plausibility of one model in comparison to another, as described in Section G4.3.

Table G13-3. Comparison of Parameter Estimates (Standard Errors) Obtained from Five Competing Statistical Models Which Regress Blood-Lead Concentration on Measures of Drip-Line Soil-Lead Concentration, Floor Dust-Lead Loading, Window Sill Dust-Lead Loading, Paint/Pica Hazard and Race.

Parameter	Parameter Estimates(Standard Errors) for the Five Statistical Models				
	Log-Linear[a]	Log-Additive	Alternative[a] Log-Additive	Active Uptake	Active/ Passive Uptake
β_0 Intercept	0.399 (0.216)	4.083 (0.308)	-0.658 (0.952)	4.409 (0.559)	4.409 (1.262)
β_1 Floor	0.058 (0.036)	0.00008 (0.00031)	0.131 (0.213)	0.00002 (0.00047)	0.00002 (0.00048)
β_2 W. Sill	0.065 (0.033)	0.00097 (0.00037)	0.280 (0.191)	0.003 (0.002)	0.003 (0.004)
β_3 Soil	0.109 (0.032)	0.00042 (0.00018)	0.609 (0.169)	0.00090 (0.00056)	0.00090 (0.00100)
β_4 Paint/Pica	0.209 (0.090)	1.688 (0.870)	1.604 (0.835)	2.741 (2.387)	2.741 (3.719)
β_5 Black	0.514 (0.079)	3.567 (0.622)	3.483 (0.604)	6.413 (3.516)	6.412 (7.491)
θ[b]				25.174 (17.194)	25.175 (76.900)
$F_{Passive}$[c]					0.000 (0.994)
^{2}	0.255	0.277	0.265	0.278	0.278
R^2	0.371	0.316	0.346	0.323	0.323
ln( )[d]	-128.623	-136.253	-132.190	-135.353	-135.375

[a] In the implementation of the log-linear and alternate log-additive models, the categorical variables Paint/Pica and Black were not log-transformed.

[b] The parameter θ appears in the Active Uptake and Active/Passive Uptake Models described in Section G3.0.

[c] The parameter $F_{Passive}$ appears in the Active/Passive Uptake Model described in Section G3.0.

[d] Ln( ) represents the log-likelihood of the observed Rochester data given each model, and can be used to assess the plausibility of one model in comparison to another, as described in Section G4.3.

Table G13-4. Comparison of Parameter Estimates (Standard Errors) Obtained from Five Competing Statistical Models Which Regress Blood-Lead Concentration on Measures of Drip-Line Soil-Lead Concentration, Floor Dust-Lead Loading, Window Sill Dust-Lead Loading and Paint/Pica Hazard.

Parameter	Parameter Estimates(Standard Errors) for the Five Statistical Models				
	Log-Linear[a,e]	Log-Additive	Alternative[a] Log-Additive	Active Uptake	Active/ Passive Uptake
β_0 Intercept	0.418 (0.240)	4.932 (0.355)	-0.528 (1.124)	5.229 (1.034)	5.229 (1.134)
β_1 Floor	0.066 (0.040)	0.00000 (0.00037)	0.371 (0.276)	0.00000 (0.00172)	0.00000 (0.00173)
β_2 W. Sill	0.087 (0.036)	0.00126 (0.00046)	0.276 (0.213)	0.012 (0.009)	0.012 (0.012)
β_3 Soil	0.114 (0.035)	0.00050 (0.00022)	0.651 (0.206)	0.003 (0.002)	0.003 (0.002)
β_4 Paint/Pica	0.248 (0.100)	2.304 (1.035)	2.139 (0.996)	10.823 (11.740)	10.824 (14.748)
θ[b]				13.365 (3.317)	13.365 (7.597)
$F_{Passive}$[c]					0.000 (0.063)
^{2}	0.316	0.342	0.323	0.327	0.327
R^2	0.217	0.153	0.190	0.198	0.198
ln( )[d]	-148.303	-155.331	-151.361	-150.393	-150.412

[a] In the implementation of the log-linear and alternate log-additive models, the categorical variable Paint/Pica was not log-transformed.

[b] The parameter θ appears in the Active Uptake and Active/Passive Uptake Models described in Section G3.0.

[c] The parameter $F_{Passive}$ appears in the Active/Passive Uptake Model described in Section G3.0.

[d] Ln( ) represents the log-likelihood of the observed Rochester data given each model, and can be used to assess the plausibility of one model in comparison to another, as described in Section G4.3.

[e] Note that the log-linear model which regresses blood-lead on floor dust-.lead loading, window sill dust-lead loading, dripline soil-lead concentration and paint/pica hazard is the unadjusted Multi-media predictive model described in Section 5 of this document.