

## **APPENDIX C**

### **METHOD TO IMPUTING HOUSEHOLD AVERAGE ENVIRONMENTAL-LEAD LEVELS FOR HOUSING UNITS IN THE NATIONAL SURVEY OF LEAD AND ALLERGENS IN HOUSING (NSLAH)**

## Method to Imputing Household Average Environmental-lead Levels for Housing Units in the National Survey of Lead and Allergens in Housing (NSLAH)

Occasionally, some of the 706 housing units included in the interim NSLAH database had no data available to calculate one or more of the following five environmental-lead parameters:

- area-weighted household average floor dust-lead loading
- area-weighted household average window sill dust-lead loading
- household average soil-lead concentration at dripline/entryway
- household average soil-lead concentration at mid-yard
- yard-wide average soil-lead concentration (taken to be the average of the previous two measures, or only one of these two measures if no data exist for the other).

In order to apply the risk analysis to the NSLAH data (specifically, the modeling analysis), it was necessary to estimate these parameters in situations where their values could not be calculated for a given housing unit due to a lack of available data (i.e., no floor dust-lead loading data, no window sill dust-lead loading data, or no soil-lead concentration data). Otherwise, those housing units having missing data, and the portion of the national housing stock represented by their sampling weights, could not be represented in the risk analysis. The method of assigning estimated data values to housing units having missing data is called *imputation*.

The imputation method applied to the interim NSLAH data was the same method used in the §403 risk analysis to impute environmental-lead levels for HUD National Survey units. This method was documented in Section 3.3.1.1 and Appendix C of the §403 risk analysis report. This method involved the following:

1. Each NSLAH housing unit was placed into one of 15 categories defined by the combination of five housing age categories (pre-1940, 1940-1959, 1960-1977, post-1977, unknown) and three categories determined by whether or not lead-based paint (LBP, defined as paint with an x-ray fluorescence measurement of at least 1.0 mg/cm<sup>2</sup>) was observed in the unit (yes, no, unknown).
2. Within the eight categories in which both the housing age group and the presence of LBP were known, the weighted averages of the first four environmental-lead parameters above were calculated across the housing units having nonmissing data (where the weights corresponded to the interim NSLAH sampling weights). Then, within a given category, if a housing unit had missing data for one of these four parameters, the weighted average for that parameter was assigned to the unit.
3. For the category in which both the housing age group and the presence of LBP were unknown, housing units having missing data for a given parameter among the first four

parameters above were assigned the weighted average for that parameter calculated across all units in the interim NSLAH database having nonmissing data for that parameter.

4. For the four categories in which the housing age group was specified but the presence of LBP was unknown, housing units having missing data for a given parameter among the first four parameters above were assigned the weighted average for that parameter calculated across units within the same housing age group (without regard to the presence of LBP) that had nonmissing data for that parameter.
5. For the two categories in which the presence of LBP was known but the housing age group was not specified, housing units having missing data for a given parameter among the first four parameters above were assigned the weighted average for that parameter calculated across units having the same indicator of LBP (without regard to housing age group) that had nonmissing data for that parameter.
6. If a housing unit had a missing value for yard-wide average soil-lead concentration (i.e., no soil-lead concentration data for any soil samples), the parameter's imputed value assigned to this unit was the arithmetic average of the unit's imputed values for average dripline/entryway soil-lead concentration and average mid-yard soil-lead concentration. (Note that if soil-lead data existed for one location but not for the other, the unit's yard-wide average equaled the average for only the location having soil-lead data.)

Table C-1 presents the weighted averages that were assigned to units having missing data as part of this imputation scheme, according to category. Note that only those weighted averages that were assigned to at least one housing unit with missing data are displayed in this table. The numbers in parentheses correspond to the numbers of housing units in the category to which the given weighted average was assigned. Only 11 of the 15 housing unit categories are included in Table C-1, as no imputations were necessary in the other four categories.

As indicated in Table C-1, the above imputation procedure was applied twice to the NSLAH data: once when making no adjustments to not-detected values, and once after replacing not-detected values with one-half of the detection limit. Both of these scenarios were considered in the data summaries and risk analysis. In both cases, the imputed values were the same in a majority of situations, and those differences which did occur between the two cases were minor.

**Table C-1. Imputed Environmental-Lead Measurements, by Housing Age Category and Presence of Lead-Based Paint (LBP)<sup>1</sup>, and Numbers of Units in the Interim NSLAH to Which Imputed Measurements Were Assigned**

Household Average Environmental-Lead Measurement	Imputed Measurement <sup>2</sup> (Number of Interim NSLAH units in which imputed measurements were assigned)										
	Pre-1940 Units		1940-1959 Units		1960-1977 Units		Post-1977 Units			Units with Housing Age Unspecified	
	LBP Present	LBP Not Present	LBP Present	LBP Not Present	LBP Present	LBP Not Present	LBP Present	LBP Not Present	LBP Presence Unknown	LBP Present	LBP Not Present
<b>No Adjustment Made to Not-Detected Values</b>											
Floor Dust-Lead Loading ( $\mu\text{g}/\text{ft}^2$ )	35.30 (1)	--	4.94 (2)	--	--	1.24 (3)	--	1.18 (1)	1.20 (1)	21.20 (1)	--
Window Sill Dust-Lead Loading ( $\mu\text{g}/\text{ft}^2$ )	449.06 (3)	15.45 (1)	144.42 (4)	94.66 (4)	--	28.95 (12)	28.81 (1)	13.99 (12)	15.62 (1)	285.64 (1)	32.73 (2)
Yard-Wide Average Soil-Lead Concentration <sup>3</sup> ( $\mu\text{g}/\text{g}$ )	710.77 (7)	176.71 (3)	276.07 (4)	242.58 (3)	161.91 (3)	52.33 (5)	--	24.85 (7)	27.78 (1)	392.05 (5)	63.84 (4)
Soil-Lead Concentration at Dripline/Entryway ( $\mu\text{g}/\text{g}$ )	1094.6 (8)	223.48 (5)	399.75 (6)	344.61 (3)	245.35 (3)	64.45 (8)	--	27.15 (8)	31.88 (1)	591.39 (5)	80.84 (4)
Soil-Lead Concentration at Mid-Yard ( $\mu\text{g}/\text{g}$ )	326.95 (8)	129.93 (3)	152.39 (7)	140.55 (3)	78.47 (4)	40.20 (8)	--	22.56 (11)	23.68 (1)	192.71 (5)	46.84 (5)
<b>Not-Detected Values Replaced by LOD/2 (i.e., one-half of the detection limit)</b>											
Floor Dust-Lead Loading ( $\mu\text{g}/\text{ft}^2$ )	35.47 (1)	--	5.19 (2)	--	--	1.72 (3)	--	1.71 (1)	1.71 (1)	21.45 (1)	--
Window Sill Dust-Lead Loading ( $\mu\text{g}/\text{ft}^2$ )	449.10 (3)	15.81 (1)	144.76 (4)	94.88 (4)	--	29.28 (12)	28.90 (1)	14.43 (12)	16.02 (1)	285.81 (1)	33.09 (2)
Yard-Wide Average Soil-Lead Concentration ( $\mu\text{g}/\text{g}$ )	710.82 (7)	176.62 (3)	276.10 (4)	242.76 (3)	162.07 (3)	52.86 (5)	--	25.73 (7)	28.57 (1)	392.15 (5)	64.43 (4)
Soil-Lead Concentration at Dripline/Entryway ( $\mu\text{g}/\text{g}$ )	1094.6 (8)	223.48 (5)	399.76 (6)	344.66 (3)	245.47 (3)	64.85 (8)	--	27.86 (8)	32.52 (1)	591.46 (5)	81.30 (4)
Soil-Lead Concentration at Mid-Yard ( $\mu\text{g}/\text{g}$ )	327.01 (8)	129.75 (3)	152.45 (7)	140.86 (3)	78.67 (4)	40.87 (8)	--	23.60 (11)	24.63 (1)	192.84 (5)	47.56 (5)

<sup>1</sup> Units with lead-based paint have a maximum observed XRF reading of at least 1.0 mg/cm<sup>2</sup> on interior or exterior painted surfaces.

<sup>2</sup> See text for details on method of determining imputed measurements.

<sup>3</sup> Imputed only when unit has no soil-lead data for either dripline/entryway or mid-yard.

