4.0 DOSE-RESPONSE ASSESSMENT

The objective of dose-response assessment was to characterize the relationship between environmental-lead exposure and the resulting adverse health effects in young children. The foundation of this characterization was the relationship between environmental-lead levels and blood-lead concentration. EPA's Integrated Exposure, Uptake, and Biokinetic (IEUBK) model and an empirical model developed for this risk assessment (Sections 4.1 and 4.2 of the §403 risk analysis report) were employed to make this characterization.

Section 4.1 of this chapter documents an additional tool, obtained since the §403 risk analysis report was published, for predicting blood-lead concentration as a function of environmental-lead levels. This tool is a regression model developed from epidemiological data collected from 12 studies and was suggested for use in the §403 risk analysis by some commenters on the §403 proposed rule. As the U.S. Department of Housing and Urban Development (HUD) sponsored the development of this model, it is referred to in this report as the "HUD Model." The goal of this model was to "estimate the contribution of lead-contaminated house dust and soil to children's blood-lead levels" (Lanphear et al., 1998). This goal is consistent with the objectives of the §403 risk assessment, and so the model merits consideration for this analysis. Section 4.1 includes documentation on the HUD model, key steps that were taken in its development, and issues that are necessary to consider when interpreting the results of the model fits.

Section 4.2 of this chapter contains a revision of the Rochester Multimedia model, introduced in Section 4.2.3 of the §403 risk analysis report, to allow the model to predict results that are more comparable to the results of the performance characteristics analysis presented in the preamble to the §403 proposed rule. The Rochester Multimedia model predicted a geometric mean blood-lead concentration as a function of average dust-lead loadings in floors and window sills, dripline soil-lead concentration, and a variable which indicates the presence of deteriorated lead-based paint and a child with paint pica tendencies. In contrast, the performance characteristics analysis in the preamble estimated risks associated with dust-lead loadings on uncarpeted floors, dust-lead loadings on window sills, yardwide average soil-lead concentration, and the percentage of painted components with deteriorated lead-based paint. Because the definitions of the data inputs were not always consistent between these two statistical approaches, their findings were not comparable. Thus, the revised model presented in Section 4.2 uses the same types of data inputs as those used for the performance characteristics analysis. Section 4.2 also documents multimedia models that omit one or more of the dust, soil, and paint input variables to obtain predicted blood-lead concentration in instances where data for one or more of these media were not available.

Section 4.3 provides additional information regarding key assumptions made in the risk characterization process: the "scaling" algorithm used to determine a post-intervention blood-lead concentration distribution that is comparable to the baseline distribution, and the issue of adjusting for measurement error when deriving the empirical model used in the §403 risk analysis.

4.1 HUD MODEL

The HUD model (Lanphear et al., 1998) was developed by a team of researchers and sponsored by HUD's Office of Lead Hazard Control. This modeling effort used data from 12 epidemiologic studies (hence its frequent reference as a "pooled analysis" model) to make statistical inferences on the contribution of lead-contaminated house dust and residential soil to children's blood-lead concentration.

The HUD model predicts a geometric mean blood-lead concentration for children aged 6-36 months (i.e., the age range of data considered from the 12 studies) as a function of exposure to specific lead levels in dust, soil, paint, and water, and as a function of other important demographic variables. Therefore, the model is used to estimate individual risks, or the risks associated with a class of children determined by specified environmental-lead levels to which they are exposed. EPA addressed minimizing individual risks when establishing "levels of concern" for lead in dust and soil within the §403 proposed rule. However, EPA was obliged to consider population-based risks within a cost-benefit analysis to establish lead hazard standards within the proposed rule.

4.1.1 Form of the HUD Model

The HUD model takes the following form:

$$
\begin{aligned}
(1)\quad Ln(PbB) = {} & 1.496 + 0.183 \cdot Ln(DustLead) + 0.01398 \cdot Ln(WaterLead) + 0.02116 \cdot Ln(ExtLead) + \\
& 0.005787 \cdot Ln(ExtLead) \cdot ExtType + 0.4802 \cdot Ln(ExtLead) \cdot ExtLoc - 0.1336 \cdot ExtType + \\
& 0.5858 \cdot ExtLoc - 0.02199 \cdot Ln(MaxXRF) + 0.03811 \cdot Ln(MaxXRF) \cdot PaintCond - 0.0808 \cdot PaintCond \\
& + 0.02126 \cdot Age - 0.001399 \cdot Age2 + 0.00007854 \cdot Age3 - 0.3932 \cdot Boston - 0.01167 \cdot Butte + \\
& 0.2027 \cdot Bcreek + 0.2392 \cdot Cpgm + 0.5383 \cdot Csoil + 0.05717 \cdot Leadville + 0.1761 \cdot Magna - \\
& 0.04209 \cdot RochLong + 0.07257 \cdot RochLID - 0.3712 \cdot Sandy + 0.1777 \cdot Midvale + 0.123 \cdot Race + \\
& 0.3175 \cdot SES1 + 0.2138 \cdot SES2 + 0.1799 \cdot SES3 + 0.1691 \cdot SES4 - 0.03233 \cdot MouthOften - \\
& 0.2454 \cdot MouthRare - 0.1397 \cdot MouthSome + 0.002649 \cdot Ln(DustLead) \cdot Age - \\
& 0.0003381 \cdot Ln(DustLead) \cdot Age2 - 0.00001281 \cdot Ln(DustLead) \cdot Age3 + \\
& 0.2212 \cdot Ln(ExtLead) \cdot MouthOften + 0.07892 \cdot Ln(ExtLead) \cdot MouthRare + \\
& 0.1663 \cdot Ln(ExtLead) \cdot MouthSome + 0.5305 \cdot Ln(WaterLead) \cdot SES1 - 0.0136 \cdot Ln(WaterLead) \cdot SES2 \\
& + 0.1033 \cdot Ln(WaterLead) \cdot SES3 - 0.09098 \cdot Ln(WaterLead) \cdot SES4 + 0.01192 \cdot Age \cdot Race - \\
& 0.01023 \cdot Age \cdot SES1 + 0.003849 \cdot Age \cdot SES2 + 0.00008468 \cdot Age \cdot SES3 - 0.01679 \cdot Age \cdot SES4 + error
\end{aligned}
$$

where

Ln(PbB) = log-transformed blood-lead concentration (µg/dL)

Ln(DustLead) = log-transformed interior (wipe) floor dust-lead loading (µg/ft^2), minus the mean of the log-transformed data used to develop the model (2.605 µg/ft^2)

Ln(WaterLead) = log-transformed water-lead concentration (ppb), minus the mean of the log-transformed data used to develop the model (0.785 ppb)

Ln(ExtLead) = log-transformed exterior-lead concentration (ppm), minus the mean of the log-transformed data used to develop the model (6.232 ppm), where the exterior sample is either soil collected at the perimeter of the foundation, soil from the child's play area, or exterior dust

ExtType = indicator of the type of exterior sample (1 = dust, 0 = soil)

ExtLoc = indicator of whether exterior sample is represented by soil at the perimeter of the house's foundation (1 = exterior sample is <u>not</u> from perimeter soil, 0 = exterior sample is from perimeter soil)

Ln(MaxXRF) = log-transformed maximum lead-paint measurement on interior surfaces (mg/cm^2, as measured by XRF), minus the mean of the log-transformed data used to develop the model (0.921 mg/cm^2)

PaintCond = indicator of paint condition (1 = damaged, 0 = undamaged)

Age = age of child (months) minus the mean age of children whose data were used to develop the model (16.3 months)

Age2 = Age^2 - (85.5 + 4.82·Age) (quadratic orthogonal polynomial)

Age3 = Age^3 - (-490.71 + 10.32·Age2 + 122.3·Age) (cubic orthogonal polynomial)

Boston, Butte, Bcreek, Cpgm, Csoil, Leadville, Magna, RochLong, RochLID, Sandy, and *Midvale* are indicators that the data come from the particular study being represented (1 = data comes from the particular study, 0 = otherwise)

Race = Race indicator (0 = white, 1 = other)

SES1 = indicator of whether the pseudo-Hollingshead measure of socioeconomic status is equal to 1 (1 = yes, 0 = no)

SES2 = indicator of whether the pseudo-Hollingshead measure of socioeconomic status is equal to 2 (1 = yes, 0 = no)

SES3 = indicator of whether the pseudo-Hollingshead measure of socioeconomic status is equal to 3 (1 = yes, 0 = no)

SES4 = indicator of whether the pseudo-Hollingshead measure of socioeconomic status is equal to 4 (1 = yes, 0 = no)

MouthOften = indicator of whether mouthing behavior occurs often in the child (1 = often, 0 = otherwise)

MouthRare = indicator of whether mouthing behavior occurs rarely in the child (1 = rarely, 0 = otherwise)

MouthSome = indicator of whether mouthing behavior occurs sometimes in the child (1 = sometimes, 0 = otherwise)

error = random error between the observed log-transformed blood-lead concentration and what is predicted by the model.

4.1.2 Development of the HUD Model

This section presents several issues on how the HUD model was developed that have a direct impact on the predicted blood-lead concentration and how this prediction should be interpreted. These issues include how studies were selected, how study effects were represented in the model, and how data were handled or adjusted prior to or during the model development exercise.

Study Selection and Potential Selection Bias

The HUD model was developed from environmental-lead, blood-lead, and demographic data from 12 studies performed over a 15-year time frame (1982-1997). These studies investigated the relationship between environmental-lead levels and children's blood-lead levels in various locations and subpopulations. Five of the studies (representing 62% of the data used to fit the model) were conducted in urban environments:

- Boston Longitudinal Study (Rabinowitz et al., 1985)
- Cincinnati Longitudinal Study (Bornschein et al., 1985b)
- Cincinnati Soil Study (Clark et al., 1991)
- Rochester Longitudinal Study (Lanphear et al., unpublished)
- Rochester Lead-in-Dust Study (Lanphear et al., 1996a,b)

The remaining seven studies (representing 38% of the data used to fit the model) were conducted in milling, mining, or smelter environments:

- Bingham Creek, Utah (1993)
- Butte, Montana (1990)
- Leadville, Colorado (1991)
- Magna, Utah (1994)
- Midvale, Utah (1989)
- Palmerton, Pennsylvania (1994)
- Sandy, Utah (1994)

According to Lanphear et al. (1998), these 12 studies were selected based on the following criteria:

- The studies had well-defined sampling protocols for blood and environmental media (particularly dust, soil, and paint).
- The studies took measures of dust-lead levels, soil-lead levels, paint-lead content (via XRF), and paint condition.
- The original data were available and could be reanalyzed.
- Dust samples were collected via wipe techniques or by the Dust Vacuum Method (DVM)
- Dust samples were taken within three months of collecting blood samples from the resident child(ren) (to address seasonal variation in blood-lead concentration).
- Children were <u>not</u> selected on the basis of having a high blood-lead concentration.

In addition, only cross-sectional data were considered (i.e., data were not considered that could reflect changes in environmental-lead levels over time).

As a result of the inclusion criteria, data for at least seven other studies considered by HUD were excluded from the model development process. These studies and the primary reasons for their exclusion (when specified within Lanphear et al., 1998) were

- UK Study (Davies et al., 1990) – dust collection method not wipe or DVM
- Boston and Baltimore segments of the EPA Urban Soil Lead Abatement Demonstration Study (USEPA, 1996a; Weitzman et al., 1993) – dust collection method not wipe or DVM
- Australian National Survey (1996) – lack of XRF paint-lead levels
- Baltimore R&M Study (USEPA, 1996c; USEPA, 1997c) – method for selecting children did not meet the criteria, and data were not available to the analysis
- Telluride, CO (1987) – reason for exclusion not given
- Trail, BC (1992) – reason for exclusion not given.

Lanphear et al. (1998) indicates that the 12 studies were not chosen to represent the entire nation or even communities like those in which the studies were conducted. In fact, these studies were conducted in communities with a recognized environmental-lead hazard, and any abatement efforts within each study targeted those hazards. Furthermore, the study effects included in the model were treated as fixed effects (i.e., they are the only studies of interest in the model-building process) rather than random effects (i.e., they are assumed to be a random sample of a larger population of studies). Thus, if the model is used to estimate risks to a broader population of children than simply those within the 12 studies, additional information is needed to determine the extent to which the pooled data used to develop the model are representative of the U.S. housing stock.

<u>Fixed vs. Random Study Effects and Interaction with the Study Effects</u>

As stated in the previous paragraph, the study effect in the HUD model is a series of fixed effects. If the study effect was assumed to be random instead of fixed (i.e., the studies can be considered a random selection of all such residential-lead exposure studies), then study-to-study variation would become a contributor to total variation in the prediction. Based on work with previous models that incorporate a random study effect, the study-to-study component is typically a major portion of total variability in the prediction. Thus, the variability associated with predictions by the HUD model is likely underestimated.

Additional underestimation in variability may result from the absence of interaction terms in the model between study effects and other environmental exposure factors. This can underestimate variability associated with inferences involving the environmental exposure factors, including the principal inferences which involve dust-lead and exterior-lead levels.

<u>Adjusting for Measurement Error in Environmental-Lead Predictor Variables</u>

The HUD model parameter estimates were determined from Simulation Extrapolation (SIMEX) methods, which attempted to quantify the theoretical relationship between blood-lead and “error-free” measures of environmental-lead. As a result, the parameter estimates were adjusted to reflect measurement error present among the environmental-lead variables. However, the goal is to predict children's blood-lead concentrations as a function of wipe dust-lead loadings and soil-lead concentrations as they would be measured in a risk assessment, not their true, "error-free" (but unobservable) values. Carroll et al. (1995) states that for prediction problems, adjusting for the effects of measurement error in predictor variables is rarely necessary. Adjusting for measurement error in these predictor variables tends to increase the values of the slope parameters associated with these variables, which in turn can inflate predicted blood-lead concentrations. Thus any predictions from the HUD model fits should be properly labeled that values of the predictor variables are assumed to be "error-free" rather than measures of environmental-lead levels taken from activities such as a risk assessment (as the predictor variables in the models used in the §403 risk analysis were assumed to represent).

Making Survey Variable Definitions Consistent Across Studies

Because different survey designs in different studies can result in different definitions for a common survey measure (e.g., SES, mouthing behavior, paint condition), which in turn can introduce considerable complication when interpreting model predictions, certain survey measures were redefined to make them more consistent across studies. Each redefinition transforms the original data values to values generated from a domain that is consistent across the studies. However, such a transformation does not remove all study-to-study differences in these values. In particular, it does not consider factors that impact how the specific study measurements were obtained and which differ from study to study, such as the use of different survey instruments and different approaches to administering the instruments.

Converting DVM Dust-Lead Loadings to Wipe-Equivalents

While it was desired to have floor dust-lead loading assuming wipe dust collection as a predictor variable in the HUD model, some of the 12 studies used DVM methods to collect dust samples. Rather than exclude data from these studies from consideration in the model development effort, a procedure was derived to convert DVM dust-lead loadings reported in these studies to wipe-equivalent loadings. This procedure used data from the Butte study to develop the following conversion equations:

- $Log_{10}(Wipe) = 0.7727 + 0.9821 \cdot Log_{10}(DVM)$ *for carpeted floors,*
- $Log_{10}(Wipe) = 0.1762 + 0.4839 \cdot Log_{10}(DVM)$ *for hard floors.*

where "Wipe" and "DVM" indicate wipe dust-lead loadings and DVM dust-lead loadings, respectively (Westat, 1998). Note that these same two equations were used to convert DVM dust-lead loadings in each study, regardless of whether the relationship between DVM and wipe dust-lead loadings differed among the studies.

The method for deriving these conversion equations included a procedure to adjust for measurement error in the DVM dust-lead loading measurements. However, the purpose of the conversion was to predict a measured wipe dust-lead loading based on a measured DVM dust-lead loading, not the (unobservable) "true" DVM dust-lead loading. Thus, some bias may have been introduced in this conversion process.

Interpreting the "Exterior-Lead" Predictor

Certain households whose data were used in the HUD model development effort did not have soil-lead data for various reasons (e.g., no bare soil). In these instances, exterior dust-lead concentration was generally measured instead. As a result, among the predictor variables in the HUD model was an indicator variable that identified whether or not an exterior-lead measurement was from dust or soil. This indicator allowed both the model intercept and the slope factor associated with exterior-lead concentration to change according to its value. However, no other consideration was made for differences in sampling and analysis methods between soil and exterior dust and the impact such differences can have on the reported lead levels. In addition, no indication was given that differences in bioavailability between soil-lead and exterior dust-lead were considered in developing the model.

When a household had soil-lead data available, data from foundation perimeter (i.e., dripline) soil were used when available; otherwise, play-area soil-lead data were used. Like the soil vs. exterior-dust issue in the previous paragraph, the HUD model includes an indicator variable that identified whether or not soil-lead levels represent dripline soil. This indicator allowed both the model intercept and the slope factor associated with exterior-lead concentration to change according to its value. However, certain study-to-study differences in collecting soil samples or obtaining a soil-lead measurement were not considered in model development, such as the depth of soil sampling, soil surface type (e.g., covered vs. bare), chemical methods for the digestion and analysis of soil samples, and soil compositing.

Handling Missing Water-Lead Measurements

While the HUD model included water-lead concentration as a predictor variable, it was necessary to impute values for this measurement during model development when data for a given household were not available. Water-lead data were unavailable for all households in two studies and up to 12% of study households in the other ten studies. In such instances, imputed measurements were randomly generated from a lognormal distribution with geometric mean equal to that observed from data for other study households (if data for other households were available) or to the community-wide average (if data for other households were not available).

Handling Data Reported Below a Detection Limit

When data values reflected measurements at or below some detection limit, the HUD model development effort replaced them with probability-based values between zero and the detection limit, where probabilities were determined from a lognormal distribution associated with data above the detection limit. According to Table 2.15 of Westat (1998), the incidence of not-detected results was high with XRF paint-lead level, and to a lesser extent, with water-lead concentration. In half of the studies, the percentage of not-detected paint-lead results ranged from 20% to 86% (with the detection limit being reported at either 0.1 or 0.7 mg/cm^2). The percentage of not-detected water-lead concentrations (i.e., results somewhere below 5 ppb) exceeded 80% in two studies. In contrast, none of the lead levels in blood, dust, or soil samples were reported below a detection limit in 10 of the 12 studies, and lead levels in no more than 9% of these samples were reported below a detection limit in the other two studies.

Selecting from Multiple Observations Within a Household

When blood-lead concentration data were available for multiple children within a household (as occurred in nine of the studies), only data for one child selected at random from the household were considered in the HUD model development. If blood-lead data existed at multiple time points for the same child (such as at 6, 18, and 24 months of age in the Boston Longitudinal study), those time points whose data met the initial inclusion criteria were identified (e.g., lead interventions did not occur between the time points), and data for the time point having dust-lead measurements taken more closely in time were selected for the model development. Similarly, when environmental-lead measurements were repeatedly taken over time for a given household, data for the time point closest to a blood-lead measurement were used.

Handling Seasonality Effects

The effect of seasonality on blood-lead concentrations was given some consideration in the modeling effort (e.g., blood and dust samples must have been collected within three months of each other for their data to be included). However, there is no effect of seasonality included in the final model. It is unclear whether seasonality was determined not to be a significant effect among the pooled data, or whether a seasonality term was intentionally left out of the model.

4.1.3 Interpreting Results of Fitting the HUD Model

The previous section discussed issues concerning the pooled study data and development of the HUD model that should be understood when using the model to estimate risks, as is done in Section 5.1.1 and Appendix F. This section addresses the interpretation of results from fitting the HUD model and, in particular, caveats associated with certain interpretations.

Individual risks vs. population-based risks

As mentioned earlier, the HUD model estimates individual risks associated with lead exposure to children aged 6-36 months. While the §403 risk analysis included individual risks analyses, which EPA used in efforts to establish levels of concern for lead in environmental media, EPA was required to employ cost-benefit analysis to select §403 hazard standards. The cost-benefit analyses used population-based risks (i.e., risks posed by childhood lead exposure to the nation as a whole) to estimate the benefit and cost associated with performing interventions and other activities in response to §403 rules.

Individual risks and population-based risks are generally not comparable. This must be understood when attempting to compare the individual risks estimated by fitting the HUD model at specified environmental-lead levels with the population-based risk estimates found in the §403 risk analysis.

Interpreting Model Parameter Estimates

The prediction parameters in the HUD model are not independent. For example, it is known that soil-lead and dust-lead measures are correlated. Therefore, it is not appropriate to interpret the parameter estimates in the HUD model (or in the models developed for the §403 risk analysis) in isolation. Using the parameter estimates to characterize a cause-and-effect relationship that is attributable to a single parameter alone, such as measuring the extent of an increase in blood-lead concentration associated with a given increase in dust-lead loading, is very problematic.

One example of how correlation among the predictor variables can influence the model parameter estimates is seen with maximum XRF paint-lead measurement. One would expect a positive correlation between maximum XRF paint-lead measurement and blood-lead concentration, and as a result, a positive slope parameter. However, the estimated slope parameter is negative (-0.022), although not significantly different from zero. The negative estimate is likely due to confounding between paint-lead measurements and other predictor variables. The likelihood of confounding increases with the number of parameters in the model.

Problems with interpreting model parameter estimates in isolation emphasizes the need to consider total exposure (i.e., prediction based on considering the joint effect of all model parameters) rather than exposure associated with a single environmental medium. In the §403 situation, protectiveness needs to be judged by recognizing that hazard standards exist for dust, soil, and paint, and that resulting actions from these multiple standards will determine the level of protection, not just the actions associated with a single standard. For example, the level of protection associated with a dust-lead loading standard of 5 $\mu g/ft^2$, without consideration of other standards, may equal that associated with a joint set of standards that involve a higher dust-lead loading standard.

Interpreting Results at Low Environmental-Lead Exposures

The HUD model and the models developed for the EPA risk analysis are "log-log" models. That is, they predict log-transformed blood-lead concentration as a linear function of log-transformed environmental-lead levels. As the log transformation "stretches out" the lower portion of the scale and contracts the upper portion of the scale, very low environmental-lead levels and blood-lead concentrations have undue influence on inferences made from the models. For example, the effect of increasing dust-lead loading from 1 to 10 $\mu g/ft^2$ is equal to the effect of increasing dust-lead loading from 10 to 100 $\mu g/ft^2$. Therefore, inferences at such low levels can be overestimated and misleading. Thus, any inferences at very low dust-lead loadings, such as 1 or 5 $\mu g/ft^2$, should be made with caution.

4.1.4 Conclusions

The following conclusions can be made on the HUD model and the comparison of risk estimates originating from this model versus those originating from models used in the §403 risk analysis:

- As the HUD model parameters associated with environmental-lead measurements in specific media have been adjusted for measurement error, the input parameters to this model are assumed "true" lead levels in these media. This can provide biased results when the model is used to predict blood-lead concentration associated with lead levels measured in current risk assessments. The Rochester multimedia model and the empirical model did not have such an adjustment incorporated.

- While the HUD model contains study effects, they are considered fixed effects and therefore allow the model to make predictions for only the group of children represented by the 12 studies. Furthermore, the study effects impact only the intercept of the model, and any study-to-study differences that may be present in other model terms (such as in environmental-lead measurements) are not represented.

- The HUD model handles "exterior-lead measurements" (e.g., soil) differently than the §403 models; the impact of such difference has not been determined.

4.2 ALTERNATIVE MULTIMEDIA MODELS FOR PREDICTING A GEOMETRIC MEAN BLOOD-LEAD CONCENTRATION BASED ON ENVIRONMENTAL-LEAD LEVELS

As discussed in the introduction to this chapter, the Rochester multimedia model, presented in Section 4.2.3 of the §403 risk analysis report, was developed using data from the Rochester Lead-in-Dust study to explain children's blood-lead concentration as a function of dust-lead loadings from floors (carpeted and uncarpeted) and window sills, dripline soil-lead concentration, and an indicator variable on the presence of deteriorated lead-based paint <u>and</u> a child with paint pica tendencies. This model was used in the risk characterization (Section 5.3 of the §403 risk analysis report) to determine the probability that a child exposed to specific levels of lead in paint, dust, and soil would have a blood-lead concentration at or above 10 µg/dL. EPA used these estimates of individual risk, as well as the findings of performance characteristics analyses detailed in Section 6.1 of this report, in proposing levels of concern for lead in dust (page 30318 in the §403 proposed rule).

The §403 proposed rule considered <u>uncarpeted</u> floors and a <u>yard-wide average</u> soil-lead concentration when proposing dust and soil standards and levels of concern. The performance characteristics analysis cited in the proposed rule considered these types of dust-lead and soil-lead measures. However, in the Rochester multimedia model, the floor dust-lead loading measure did not limit the type of floor surface to uncarpeted floors, the soil-lead measure represented only dripline soil, and the paint/pica indicator variable was different from the paint measure used in the performance characteristics analysis (the percentage of tested components in the home with deteriorated lead-based paint). For these reasons, it was difficult to compare estimates of individual risks based on this model to results obtained from the performance characteristics analyses. Thus, it was desired to fit an alternative multimedia model (cited as "Model A" in this section) that replaced the floor dust-lead and soil-lead predictor variables used in the Rochester multimedia model with uncarpeted floor dust-lead loading and yard-wide average soil-lead concentration, respectively, and replaced the paint/pica indicator variable with a measure of the percentage of tested components containing lead-based paint.

While a household risk assessment for lead-based paint hazards is expected, at a minimum, to characterize lead levels in floor dust and to identify the extent of deteriorated lead-based paint, it is possible that some risk assessments may not measure lead levels in soil or window sill dust. Therefore, to investigate how individual risks would be characterized in these types of risk assessments, two alternative multimedia models were fitted that were reduced versions of Model A. One model excluded soil-lead concentration as a predictor variable ("Model B"), and the other model excluded both soil-lead concentration and window sill dust-lead loading as predictor variables ("Model C").

The three alternative multimedia models were fitted using data from the Rochester Lead-in-Dust study, using the same approach used to fit the Rochester multimedia model in the §403 risk analysis. The models were log-linear in nature, where the dust-lead and soil-lead measures were log-transformed, and the models predicted a log-transformed blood-lead concentration. For example, Model A took the following form:

$$\log(PbB) = \hat{a}_0 + \hat{a}_1*\log(PbF) + \hat{a}_2*\log(PbW) + \hat{a}_3*\log(PbS) + \hat{a}_4*PbP$$

where PbB represents blood-lead concentration (µg/dL), PbF represents household average dust-lead loading for uncarpeted floors (µg/ft^2), PbW represents household average dust-lead loading for window sills (µg/ft^2), PbS represents yard-wide average soil-lead concentration (µg/g), and PbP represents the larger (between the interior and exterior of the housing unit) of the percentages of tested components containing deteriorated lead-based paint. As with the Rochester multimedia model, ordinary least squares regression methods were used to fit the models to the Rochester data.

Table 4-1 presents the estimates of the model parameters for each of the three alternative multimedia models. Note that the model fits were based on different numbers of housing units, as eliminating certain predictor variables from the above model resulted in more housing units that had all necessary data available for fitting the model. An investigation of model diagnostics showed that the extent of collinearity among the predictor variables in these models was low. Generally, the slope estimates associated with the paint variable were very low, and except for

Table 4-1. Parameter Estimates and Associated Standard Errors for the Three Alternative Multimedia Models Fitted to Rochester Study Data to Predict Log-Transformed Blood-Lead Concentration[1]

Parameter	Predictor Variable	Parameter Estimate (Standard Error)		
		Model A	Model B	Model C
$\$_0$	Intercept	0.331 (0.263)	0.899 (0.183)	1.337 (0.122)
$\$_1$	log (PbF): Area-Weighted Arithmetic Mean (Wipe) Dust-Lead Loading from Uncarpeted Floors	0.114 (0.049)	0.130 (0.048)	0.140 (0.043)
$\$_2$	log (PbW): Area-Weighted Arithmetic Mean (Wipe) Dust-Lead Loading from Window Sills	0.082 (0.037)	0.101 (0.035)	--
$\$_3$	log (PbS): Yardwide Average Soil-Lead Concentration[2] (fine soil fraction)	0.115 (0.040)	--	--
$\$_4$	PbP: The larger of the following two percentages: % of interior tested surfaces that contain deteriorated LBP, and % of exterior tested surfaces that contain deteriorated LBP[3]	0.001 (0.002)	0.002 (0.002)	0.004 (0.002)
R^2	Coefficient of Determination	20.03%	15.62%	11.38%
F_{Error}	Error	0.561	0.572	0.580
n	# data points included in model fitting	177	188	196

"--" indicates that the variable is not included in the model as a predictor. The models are log-linear in nature.

[1] One housing unit (cid=01689) had an uncarpeted floor dust-lead loading measurement of 18130.0 $\mu g/ft^2$ (only one uncarpeted floor wipe sample was collected in this unit). This data value was omitted when fitting the above models as it was highly influential and led to a noticeable reduction in the estimate of $\$_1$.

[2] Yardwide soil-lead concentration at a given housing unit was calculated as the unweighted arithmetic average of dripline and play area soil-lead concentrations. If one or the other is missing (but both are not missing), yardwide concentration was set equal to the non-missing value. If both are missing, yardwide concentration is missing.

[3] If one or the other of these two percentages is missing (but both are not missing), the value of this variable is set equal to the non-missing value. If both are missing, the value is missing.

Model C, were not significantly different from zero at the 0.05 level. See footnotes to this table for additional details on the model fits.

Other alternative multimedia models were considered when initially developing the Rochester multimedia model. These models, and information used to evaluate these models in selecting the final version used in the §403 risk analysis, were presented in Appendix G of the §403 risk analysis report.

4.3 SUPPLEMENTAL INFORMATION ON MODEL-BASED APPROACHES IN THE §403 RISK ANALYSIS

Some comments on the §403 proposed rule addressed issues concerning approaches taken in the §403 risk analysis in which statistical models were used to predict a post-intervention distribution of blood-lead concentration, and therefore, a means of assessing how health risks associated with lead-based paint hazards would change as a result of implementing the §403 rule. The types of models used in this analysis and the approach to characterize post-intervention health risks were presented in Chapters 4 and 6 of the §403 risk analysis report. In addition, technical details on these models and approaches were provided in appendices to the report. However, certain comments on the §403 proposed rule indicated that this information may not have been provided in sufficient detail or would have benefitted from additional clarity. In particular, two issues in question involved how the §403 risk analysis estimated a post-intervention blood-lead concentration distribution that was comparable to the baseline distribution that was characterized by data from Phase 2 of NHANES III, and how the empirical model (Section 4.2 of the §403 risk analysis report) account for measurement error issues associated with the predictor variables. The following subsections provide additional information on these two issues, specifically geared toward addressing the specific areas raised by selected public comments.

4.3.1 The "Scaling" Algorithm Used to Determine a Post-Intervention Blood-Lead Concentration Distribution

In the §403 risk analysis, EPA used data from Phase 2 of NHANES III (collected from 1991-1994) as the basis for the baseline ("pre-§403 rule") characterization of children's blood-lead concentration in the U.S. housing stock. As discussed in Section 3.5 of the §403 risk analysis report, EPA took this approach because these data were considered the best available data (as well as the most recent data) on blood-lead measures, the data consisted of actual blood-lead measurements from a nationally-representative survey, and it was preferred (and considered more defensible) to use such data to characterize the baseline distribution rather than data generated from statistical prediction models. However, because the "post-§403 rule" time period has not yet occurred, and it was desired to compare the blood-lead distribution in this time period to the baseline blood-lead distribution, it was necessary to use statistical prediction models to generate how the baseline distribution would change following interventions performed as a result of the §403 rule.

When estimating the blood-lead concentration and health effect endpoints used in the §403 risk analysis, it was assumed that the national distribution of blood-lead concentration was lognormally distributed. Initial investigations into the weighted NHANES III blood-lead concentration data used in the §403 risk analysis (Figure 5-3 of the §403 risk analysis report) suggested that this was a satisfactory assumption. The lognormal distribution is characterized by the geometric mean and geometric standard deviation (GSD) of the data (or, equivalently, the exponentiated mean and exponentiated standard deviation of the log-transformed data). Once the geometric mean and GSD were calculated from the weighted NHANES III data, the additional assumption of lognormality was used to obtain baseline estimates of the health effect and blood-lead concentration endpoints considered in the §403 risk analysis (e.g., probability that a child's blood-lead level was at or above 10 µg/dL). These estimates were presented in Table 5-1 of the §403 risk analysis report.

To estimate how the blood-lead distribution changed between the "pre-§403 rule" and "post-§403 rule" environments (based on a given set of candidate §403 standards and on assumed changes in environmental-lead levels resulting from implementing the §403 rule), model-based estimates of the blood-lead distribution were made for both environments. For reasons explained in the §403 risk analysis report, EPA chose to characterize both the pre-§403 and post-§403 blood-lead distributions twice, with the first characterization using the IEUBK model and the second characterization using the empirical model. Each characterization involved fitting the given model to environmental-lead data separately for each home in the HUD National Survey and weighting each prediction appropriately to represent a given proportion of the nation's children. Therefore, although the HUD National Survey homes do not represent a random sample of homes in the national housing stock, and therefore, the set of predicted blood-lead concentrations do not themselves represent a random sample of blood-lead levels in the nation's children, the fact that each prediction is weighted appropriately to represent a given proportion of the nation's children allows the total set of predictions to be a good estimate of the national blood-lead distribution, in either a pre-§403 or post-§403 environment.

Appendix E2 of the §403 risk analysis report discusses how model-predicted blood-lead concentrations generated for the HUD National Survey homes are used to estimate the geometric mean and GSD associated with the national blood-lead distribution. Recall that for a given HUD National Survey home, the model-predicted blood-lead level represents a geometric mean of children whose exposure is characterized by the environmental-lead levels in that home. Therefore, the estimated GSD of the national blood-lead distribution is characterized not only by the variability among the predicted blood-lead levels, but also by the assumed variability in blood-lead levels among individual children exposed to the same environmental-lead levels.

Under a given model (i.e., either the IEUBK or empirical model), the "scaling" algorithm (Appendix F1 of the risk analysis report) involves calculating the proportional change in the geometric mean and GSD of the model-predicted blood-lead distribution from the "pre-§403 rule" to the "post-§403 rule" environment. Then, the same proportional change in both statistics was applied to the geometric mean and GSD of the baseline distribution determined from the NHANES III data:

$$GM_{post\text{-}403} = GM_{baseline} * (GM_{model\text{-}based\ post\text{-}403} / GM_{model\text{-}based\ pre\text{-}403})$$
$$GSD_{post\text{-}403} = GSD_{baseline} * (GSD_{model\text{-}based\ post\text{-}403} / GSD_{model\text{-}based\ pre\text{-}403})$$

The resulting geometric mean and GSD, along with an assumption that lognormality still holds, characterized a blood-lead concentration that represented the post-§403 environment and that was considered comparable to the baseline distribution.

This type of algorithm was necessary due to the difficulties associated with comparing a model-based post-§403 blood-lead distribution directly with a baseline distribution that was characterized from observed (NHANES III) data. While the empirical model was calibrated so that its estimate for the baseline national geometric mean blood-lead concentration for children aged 1-2 years (as obtained within the approach taken in the §403 risk analysis) equaled the NHANES III Phase 2 estimate (although the predicted GSD was not similarly calibrated), such a calibration was not possible for the IEUBK model. As a result, using the HUD National Survey data as input, these models could not both predict the same national estimate of the geometric mean blood-lead concentration as Phase 2 of NHANES III. In addition, the empirical model could not be developed based on data from a national survey that measured both blood-lead and environmental-lead levels, which would have facilitated direct comparisons of the predicted blood-lead distribution between pre-§403 and post-§403 environments. Therefore, the “post-§403 rule” blood-lead distributions predicted by the two models had some inconsistency with the baseline distribution estimated from the NHANES III data, making direct comparisons problematic. For example, if a model underestimates the geometric mean, the benefits associated with the §403 rule could be overestimated, while if a model overestimates the geometric mean, this could result in estimates of negative benefits. Therefore, the “scaling” algorithm used in the risk analysis used the models to predict the change in the geometric mean and GSD that occurs from a pre- to post-§403 rule environment, then applied this same change to the baseline distribution.

Note that the scaling algorithm does not require that the two model-based blood-lead distributions (pre-§403 and post-§403) be independent of each other. In fact, the two distributions are dependent, because the post-§403 environmental-lead levels, used to predict post-§403 blood-lead levels, are dependent on the pre-§403 levels. Only the geometric mean and GSD of these two distributions are necessary to characterize, and they are used simply to estimate the proportional change in the geometric mean and GSD of the national blood-lead distribution between pre-§403 and post-§403 conditions.

While the geometric mean and GSD are scaled separately, one change is not necessarily independent of the other. For example, if the pre-§403 geometric mean and GSD both have high values, they are both likely to be reduced at a greater rate than at lower values. The approach was kept as simple as possible while retaining scientific defensibility, in order that it be easily applied during risk characterization and in the economic analysis.

In the peer review of the §403 risk analysis report, EPA specifically asked the peer reviewers to comment on whether the scaling procedure was scientifically defensible in general, and in particular, whether it was relevant in the situation where the environmental-lead data (from the HUD National Survey) and the blood-lead data (from NHANES III) were collected at different periods in time. None of the peer reviewers specifically criticized the scaling algorithm. Furthermore, the Science Advisory Board reviewing the §403 risk analysis stated that, in general, the approach was scientifically defensible as presented, and specifically, the multi-step approach was warranted due to the need to use various datasets of differing sources and representing different time periods to make the characterization.

See Section 6.4.4 below for an alternative approach to applying this scaling algorithm where the probability of a child's blood-lead concentration exceeding 10 µg/dL is scaled rather than the GSD.

4.3.2 Adjusting the Empirical Model Parameter Estimates to Reflect Measurement Error

The approach to the measurement error adjustment, discussed in Section 4.2.4 of the §403 risk analysis report, attempted to correct for the fact that while the empirical model was developed using data from the Rochester Lead-in-Dust study, it was used to predict a (pre-403) geometric mean blood-lead concentration assuming that the data input to the model originated from the HUD National Survey. Because the Rochester study and the HUD National Survey used different sampling schemes involving different collection devices and instruments, as well as different analytical methods, and because the ranges of observed environmental-lead levels differed between the two studies, it was necessary to adjust the model parameter estimates to reflect these differences prior to allowing the model to accept HUD National Survey environmental-lead data as input to the prediction.

Note that the measurement error adjustment made to the empirical model was not to address the more-standard "errors in variables" issue (Carroll et al., 1995) which attempts to take into account that a value input to the model represents a measurement subject to error, rather than a "true" value. As discussed in Section 4.2.4 and Section G4.2 (Appendix G) of the §403 risk analysis report, the empirical model was not intended to be used in the risk analysis as a dose-response model, which would have required the predictor variables to reflect actual exposures. Instead, the model assumed that its input environmental-lead information reflected measurements that would have been made as a result of a risk assessment within a home. Therefore, adjusting the model for the fact that its inputs reflect measured rather than actual lead levels was considered inappropriate for this analysis. This decision in the type of application that is represented by the §403 risk analysis has been concurred upon in the published literature (e.g., Carroll and Galindo, 1999).

When using the empirical model to predict a post-403 geometric mean blood-lead concentration, some of the HUD National Survey dust-lead and soil-lead data (i.e., those data for homes that exceed the candidate 403 standards) were modified to reflect the impact of performing interventions in response to the 403 rule on these measured data values (see Table 6-2 of the §403 risk analysis report), then the model is fitted to the modified data. These modified data are still considered to be measured lead levels, rather than actual (or "true") lead levels. However, the modified data values must first be transformed to represent measurements that would have made under the methods used in the HUD National Survey. For example, the assumed post-intervention floor wipe dust-lead loading of 40 µg/ft^2 must be converted to a Blue Nozzle vacuum-equivalent loading prior to using it as input to the empirical model.