Study of HUDs Risk Assessment Methodology in Three U.S. Communities

Final Report

Prepared for:
The U.S. Department of Housing and Urban Development
Office of Healthy Homes and Lead Hazard Control

By
The National Center for Healthy Housing

Volume I
Main Report and
Appendix A
January 24, 2003
Revised: June 30, 2006

Study of HUDs
Risk Assessment
Methodology in
Three U.S. Communities

Final Report

Prepared for:
The U.S. Department of Housing and Urban Development
Office of Healthy Homes and Lead Hazard Control

By
The National Center for Healthy Housing

Volume II
Appendices B-E
January 24, 2003
Final revision: June 30, 2006

VOLUME II: APPENDICES B-E

Appendix B: Enrollment Procedures for Risk Assessment Study and Study Protocols

Appendix C: Technical Appendix

Appendix D: Extended Report Tables

Authors

Warren Galke, PhD, Principal Investigator

Sherry Dixon, PhD, Statistician

Pat McLaine, RN, MPH, Program Manager

Jonathan Wilson, MPP, Project Manager/Lead Author

HUD General Technical Representative

Peter Ashley, PhD

Contributors

Baltimore County (MD) Department of Health

Maryland Department of Health and Mental Hygiene

Milwaukee (WI) Health Department

New York City (NY) Department of Public Health

Acknowledgements

Deborah Holtz, RN, MPH, MSN
Jennifer Steciak, RN, MPH
Terri Abernathy, RN, BSN
Judy Rubin, MD

Arc Environmental, Inc.
ATC Associates, Inc.
Schauer & Associates, Inc.
Tim Connor
Jim McCabe
and
Sandy Roda, University of Cincinnati

Disclaimer

This document was prepared under grant #MDLR0021-97 from the U.S. Department of Housing and Urban Development (HUD). Further funding was provided to NCHH under Task Specification #23 of Contract C-OPC-21357 between HUD and Battelle Memorial Institute. Neither the United States Government nor any of its employees makes any warranty, expressed or implied, or assumes any legal liability for any third party's use of, or the results of such use of, any information, product, or process discussed in this document. Mention or illustration of company or trade names or of commercial products does not constitute endorsement by the United States Government.

EXECUTIVE SUMMARY

In 1995 the U.S. Department of Housing and Urban Development (HUD) issued *Guidelines for the Evaluation and Control of Lead-Based Paint in Housing*. The *Guidelines* were written to provide detailed, comprehensive technical information on how to identify lead-based paint hazards in housing and how to control hazards safely and efficiently. Two protocols were defined in the *Guidelines* to facilitate the identification of housing that needed to be evaluated and possibly treated. These were the risk assessment protocol and the lead hazard screen protocol. When the *Guidelines* were released, there was a strong consensus among professionals that these protocols represented the best expert judgment available but there was also recognition that further research to validate the protocols was necessary.

In July 1995, one month after the *Guidelines* were published, the Federal Task Force on Lead-Based Paint Hazard Reduction and Financing released a report that identified research into the utility of the protocols as being a key topic for investigation. On November 27, 1996, the HUD Office of Lead Hazard Control issued a Notice of Funding Availability (NOFA) indicating that such research was a priority for its research program. In 1997, HUD awarded the National Center for Healthy Housing ((NCHH) formerly, the National Center for Lead-Safe Housing) a grant to assess HUD's risk assessment and lead hazard screen protocols found in the 1995 *Guidelines*. In 2001, EPA released regulations that changed the numeric standards for dust lead and soil lead hazards. The field test of the protocols presented in this report (National Risk Assessment Study) address the effectiveness of risk assessment and screening protocols using both the 1995 and 2001 standards.

Purpose

The primary purpose of this study was to assess under what conditions HUD's risk assessment and lead hazard screening protocols are accurate predictors of children's lead exposure. The study attempted to identify ways to improve the accuracy of the protocols. NCHH conducted a detailed, multi-media environmental assessment of residential lead in a variety of housing and linked those results to children's blood lead levels. The resulting data set served as a test bed for a number of statistical analyses that address many of the key issues regarding the identification of housing that contributes to childhood lead poisoning.

The study had the following goals:

1. To assess the ability of the current and original HUD/EPA risk assessment protocols to predict dwelling units that are likely to house children having elevated blood lead levels, and assess the effect of modifying the protocols.
2. To assess the ability of the current and original HUD/EPA lead hazard screening protocols to predict the need for risk assessments, to predict dwelling units that are likely to house children having elevated blood lead levels, and assess the effect of modifying the protocols.

3. To describe the contribution of friction and impact surfaces to floor dust lead loadings.
4. To assess the ability of the current HUD paint film quality classification system to predict rooms and dwellings that are likely to have elevated dust lead loadings.
5. To estimate the effect of dust lead measurement error on dust lead loadings.

Study Design

The National Risk Assessment Study was conducted in three locations: Baltimore County, Maryland; Milwaukee, Wisconsin; and New York City, New York. Housing in the latter two locations was chosen to represent older housing (pre-1950) while housing in Baltimore County was limited to that constructed between 1950 and 1978. The study population consisted of dwelling units housing a child, one to three years of age, who lived at the residence for at least six months prior to enrollment.

In Milwaukee and New York City, children's blood lead results, as reported by local blood lead registries and participating clinics, were used to identify dwellings eligible for the study. A case-control methodology was used in which dwellings were stratified by outcomes: half of the dwellings enrolled housed an EBL child (≥10 ug/dL) and the other half housed a child with a non-elevated blood lead level. Baltimore County did not have an accessible blood lead registry nor a blood lead screening rate sufficient to identify enough children for a four to six- month case-control study. NCHH elected to use a cross-sectional study design at this site. Potential study subjects were identified by the County based on a match of birth records and age of housing data from the tax assessor's office. Blood lead samples were drawn concurrently with the environmental sampling, so the population could not be selected on the basis of blood lead level. A local Institutional Review Board at each site approved the study design, protocols and forms.

A comprehensive set of environmental tests were taken in each home, including a visual inspection, XRF inspection, dust wipes, paint chips, soil and water samples. The environmental testing was completed in the home soon after the blood tests were reported so that it occurred either prior to or concurrent with the family receiving information on the benefits of lead-specific cleaning, to reduce the likelihood of cleaning prior to the environmental testing. Blood lead levels were collected or reported from one eligible child in the family, and a family interview was administered. The tests occurred within three weeks of each other and all the data were collected within one five-month summer "season" to reduce confounding factors. The original enrollment plan targeted 75 pre-1950 units in both New York City and Milwaukee, and 100 units built between 1950 and 1978 in Baltimore County.

From June to October 1998, certified risk assessors conducted comprehensive risk assessments/paint inspections in two hundred fifty-four dwellings. In Milwaukee and New York City, the recruitment goal of the case-control study design was achieved with 153 enrolled dwellings housing approximately equal numbers of non-EBL children (<10 μg/dL) and EBL children (≥10 μg/dL).

Table ES-1: Number of Dwellings Enrolled by Site and Elevated Blood Status

Site	Non-EBL Child Present (< 10 μg/dL)	EBL Child Present (≥ 10 μg/dL)	All Dwellings
Baltimore County	99	1	100
Milwaukee	42	37	80*
New York City	35	39	74
Total	176	77	254

*One unit was enrolled and tested based on the verbal report of a blood lead test, but the blood lead result was never confirmed by electronic report.

In Baltimore County, where the limited screening data precluded the use of a case-control design, the cross-sectional approach resulted in a sample dominated by dwellings that housed children with blood lead levels below 10 μg/dL. Since the study population was made up solely of dwellings that were built after 1950, and had a largely White population, the results are consistent with the CDC's National Health and Nutrition Examination survey (NHANES) for 1991-1994. The NHANES survey estimated that 1.4% of White children living in housing built between 1946 and 1973 would have a blood lead level above 10 μg/dL.

Summary of Analyses and Results

Goals 1 & 2: To assess the ability of the HUD/EPA risk assessment and screening protocols to predict dwelling units that are likely to house children having elevated blood lead levels, and assess the effect of modifying the protocols.

A standard method to assess the accuracy of a diagnostic test is to examine the performance characteristics of the test, using four probability measures: sensitivity, specificity, positive predictive value, and negative predictive value. All four terms are defined in the report; for this summary, sensitivity and specificity are defined below:

- *Sensitivity (or True Positive Rate)*: Probability that a dwelling unit fails an environmental assessment given that there is a resident child with an elevated blood lead concentration.
- *Specificity (or True Negative Rate)*: Probability that a dwelling unit passes an environmental assessment given that a resident child does not have an elevated blood lead concentration.

The analyses that were conducted also included a statistical test of independence between the environmental assessment result (pass/fail) and the presence or absence of a child with an elevated blood lead level. A result with a p-value less than 0.05 indicated that the environmental assessment result did not predict the child’s blood lead status.

Although the original intent of the study was to combine information from the three study sites, it proved to be inappropriate to do so. Substantial differences in both blood and environmental lead levels were found across all sites. Because only one of the 100 children enrolled in Baltimore County had an elevated blood lead level, these results could not be used to assess the effectiveness of the environmental testing protocols. As presented below, even though Milwaukee and New York City had similar aged (pre-1950) housing in the study, the environmental lead levels were very different at the two sites.

Results:

- **This study supports the premise that environmental lead results can be used to identify homes where children are likely to have elevated blood lead levels.** Analyses described below suggest that an environmental lead test can be a fairly predictive tool by maintaining the current standards but dropping window sill dust tests and assessments of paint. Further study may conclude that changes to the current standards could further improve the risk assessment protocols.
- However, **neither the current risk assessment protocols nor the screening protocols were significant predictors of the blood lead status** (< or ≥ 10 µg/dl) of a child in the dwelling.
- **Housing units in New York and Milwaukee had significantly different environmental lead levels,** although the sites had a similar distribution of children with and without elevated blood lead levels. Only water lead levels were similar at the two sites (Table ES-2).
    - **The arithmetic mean and maximum floor dust lead loading and the perimeter soil lead concentration for a dwelling were significantly different in the dwellings with and without an enrolled child with an elevated blood lead level *in Milwaukee*.** Surprisingly, neither window sill nor trough lead loadings were significantly related to blood lead status. Given the observed relationships between certain environmental lead media and blood lead levels, the home environment was assumed to be a primary source of lead exposure in Milwaukee.
    - **No environmental lead measures were significant predictors of blood lead status *in New York*.** In fact, window sill and window trough dust lead loadings, number of surfaces with non-intact interior lead-based paint and play area soil lead went in the “wrong” direction in New York, with lead levels lower in homes with children with elevated blood lead levels. Although the children and homes in New York City were enrolled under the same study design as in Milwaukee, the home environment did not appear to be the primary source of lead exposure in New York. Further analysis of data collected from household questionnaires failed to identify likely sources of the children’s elevated blood lead status.

Table ES-2: Descriptive Statistics of Environmental Lead Media by Blood Lead Outcome (EBL/Non-EBL) and Site

Statistic	Site	Number of Units	Number w/EBLs	Lead Levels (Geometric Mean)[1]	
				EBL Homes	Non-EBL Homes
Floor Dust Lead (max) (μg/ft²)	ML	64	31	45	23
	NY	69	36	8	7
Floor Dust Lead (mean) (μg/ft²)	ML	64	31	24	12
	NY	69	36	4	4
Sill Dust Lead (max) (μg/ft²)	ML	62	33	459	355
	NY	63	32	36	52
Sill Dust Lead (mean) (μg/ft²)	ML	62	30	299	247
	NY	63	32	28	43
Trough Dust Lead (max) (μg/ft²)	ML	59	27	6,749	5,171
	NY	55	28	239	422
Trough Dust Lead (mean) (μg/ft²)	ML	59	31	9,601	6,895
	NY	55	28	282	483
Perimeter Soil Lead (ppm)	ML	56	30	2,918	1,298
	NY	17	32	965	457
Play Area Soil Lead (ppm)	ML	25	14	287	261
	NY	4	3	773	948
Water Lead (first draw) (ppb)	NY	64	31	3	3
	ML	69	36	4	3
Number of LBP Surfaces-Non-Intact (Exterior)	ML	64	31	6	8
	NY	69	36	1	1
Number of LBP Surfaces-Non-Intact (Interior)	ML	64	31	18	14
	NY	69	36	4	7

[1] For Number of LBP Surfaces-Non-Intact, the arithmetic mean values are presented and tested instead of the geometric mean values.

Table ES-3: Environmental Lead Media and Standards Examined

Media	Standards Examined					
Floor Dust Lead (mean) (μg/ft²)	None	10	15	25	40	100
Sill Dust Lead (mean) (μg/ft²)	None	125	250	500		
Trough Dust Lead (mean) (μg/ft²)	None	800	5,000	10,000		
Perimeter Soil Lead (ppm)	None	400	1,200	2,000	5,000	
Play Area Soil Lead (ppm)	None	400	(1,200 was tested but no sample was above this level)			
Water Lead (first draw) (ppb)	None	5	10	15		
Number of LBP Surfaces-Non-Intact (Exterior)	None	1	5	10		
Number of LBP Surfaces-Non-Intact (Interior)	None	1	5	10		

Table ES-4: Standards for Optimal Protocols in Milwaukee

Protocol Media and Standards				Sensitivity (%)	Specificity (%)	P-value
Mean Floor Dust Pb ($\mu g/ft^2$)	Perimeter Soil Pb (ppm)	Play Area Soil Pb (ppm)	Water Pb (ppb)			
10	2,000	400	-	*100*	*36*	*<0.001*
10	2,000	-	-	*97*	*39*	*0.001*
10	-	400	-	*94*	*45*	*<0.001*
10	-	-	-	*90*	*48*	*0.001*
15	5,000	400	-	*77*	*58*	*0.006*
15	-	400	-	*77*	*58*	*0.006*
15	5,000	-	-	*84*	*55*	*0.002*
100	2,000	400	10	*77*	*58*	*0.006*
-	2,000	400	10	*74*	*61*	*0.006*
100	2,000	400	-	*74*	*61*	*0.006*
-	2,000	400	-	*71*	*64*	*0.007*
100	2,000	-	10	*71*	*64*	*0.007*
-	2,000	-	10	*68*	*67*	*0.012*
100	2,000	-	-	*68*	*67*	*0.012*
-	2,000	-	-	*65*	*70*	*0.012*
25	5,000	-	10	*61*	*73*	*0.011*
25	5,000	-	-	*58*	*76*	*0.010*
25	-	-	10	*48*	*79*	*0.035*
25	-	-	-	*45*	*82*	*0.030*
40	5,000	-	-	*39*	*85*	*0.048*

Alternative risk assessment protocols were tested using the data from Milwaukee. All permutations of the environmental lead media and standards listed in Table ES-3 were used as possible predictors of blood lead status (< or ≥ 10 μg/dl). Of the 92,190 protocols examined, 20 protocols were identified that were significant predictors of the blood lead status and optimized the performance characteristics (Table ES-4). Certain factors emerged from the results:

- **Floor dust lead loadings and perimeter soil lead concentrations were the two exposure sources most likely to be included** in the alternative protocols. These findings reinforce the earlier findings that these media were most predictive of the presence or absence of a child with an elevated blood lead level.
- **The optimal protocols included the complete range of mean floor dust lead loading standards tested (10-100 $\mu g/ft^2$).** They also included the higher levels of perimeter soil lead concentrations tested (2,000 and 5,000 ppm).
- **Some of the optimal protocols included play area soil lead (400 ppm) and water lead (10 ppb).** While the play area level matches the current standard, the water lead level is 5 ppb lower than the current action level.
- **Window sill and window trough dust lead and frequency of interior and exterior non-intact lead-based paint were not elements of the alternative protocols.** These results match the earlier findings that these media were not

predictive of homes in this study with or without a child with an elevated blood lead level.

Further analyses of data from Milwaukee explored optimal floor sampling locations:

- **The choice of floor sampling locations (Living Room, Kitchen, Bedroom, Bath and Unit Entry) and combination of locations had little difference on the ability to assess risk.** Almost all combinations of floor sampling locations were highly associated with the blood lead outcomes.
- **Floor samples taken from either *the room entry* or *central part of the floor* were generally more predictive of blood lead status** than those taken from under the window or a perimeter location.
- Although the HUD Guidelines recommend that risk assessors interview families to identify a child's "play room", **there was little difference between the predictive power of floor dust lead loadings from the "play room" versus the living room on blood lead status.** In fact, the p-values for the living room floor samples were equal or better to the play area floor samples suggesting that identifying the "play room" may not be necessary.
- **Although the choice of floor sampling locations do not appear to make a difference on the predictive power of the mean floor dust lead loadings, they may have an impact on selecting an optimal standard.** For example, the Unit Entry floor dust lead loadings were about twice as high as the interior floor dust lead loadings, so a mean floor sample result including the Unit Entry would perform differently against a given standard than a floor sample result without the Unit Entry.

Goal 3: To describe the contribution of friction and impact surfaces to floor dust lead loadings.

Risk assessors observed and recorded rubbing and/or binding on all painted doors and windows in the study. This information was included in statistical models to assess the influence of friction and impact surfaces on floor and window sill dust lead loadings.

The possible pathways of lead that are accounted for in the model included:

1. Window friction, window paint condition and window paint lead
2. Door friction, door paint condition and door paint lead
3. Lead paint (and condition) of the room
4. Exterior Lead Sources (Soil lead, other point sources)
5. Blow-in from the exterior
6. Track-in from the exterior

Results:

- **Assuming that window or door friction does produce dust lead, the results indicate that floor sampling would not be a good measure of rubbing or binding.** The interaction between the observation of rubbing/binding on doors and door paint lead and the interaction between the observation of rubbing/binding on windows and window paint lead were *not* significantly related to the floor dust lead loadings.

- **The analysis offers support to the hypothesis that window friction is a significant source of window sill dust lead even when window paint is intact.**
- **Dust lead loadings were higher on window sills where rubbing or binding was identified or window paint was not intact and dust lead loadings on those windows increased with the levels of paint lead.**

Goal 4: To assess the ability of the current HUD paint film quality classification system to predict rooms and dwellings that are likely to have elevated dust lead loadings.

When the grant for this study was awarded, HUD defined paint lead hazards as any lead-based paint in poor condition (Table ES-5). Since then, HUD and EPA issued regulations stating that all non-intact lead-based paint is a hazard. The findings for both definitions of paint deterioration are presented in the report.

Table ES-5: Categories of Paint Film Quality (HUD Guidelines Table 5.3)

	Total Area of Deteriorated Paint on Each Component		
Type of Building Component	Intact	Fair	Poor
Exterior components with large surface areas.	Entire surface is intact	Less than or equal to 10 square feet	More than 10 square feet
Interior components with large surface areas (walls, ceilings, floors, doors)	Entire surface is intact	Less than or equal to 2 square feet	More than 2 square feet
Interior and exterior components with small surface areas (window sills, baseboards, soffits, trim)	Entire surface is intact	Less than or equal to 10 percent of the total surface area of the component	More than 10 percent of the total surface area of the component

Results:

In Milwaukee,

- ***Non-intact* lead-based paint (LBP), but not *poor* LBP was a significant predictor of *floor dust lead loading*.**
- However, **the presence of *poor* LBP was a significant predictor of *blood lead status*, but not *non-intact* LBP.**
- When alternative numbers of LBP surfaces in poor condition were considered (i.e., 1, 2, 5, 10, 20 or 30), one or more LBP surfaces in poor condition had the greatest effect on the odds of having an elevated blood lead level. **A dwelling with at least one surface with poor LBP was 126% more likely to house an EBL child than a dwelling with no poor LBP.**

In Baltimore County and New York City,

- No measure of deteriorated LBP was a significant predictor of floor dust lead loading or blood lead status.

Across all three sites.

- **The results indicated that concerns about field implementation should not be a factor when determining the best method to identify deteriorated lead-based paint.** Pairs of risk assessors using the 3-level system (intact, fair, poor) to assess the condition of paint had a level of concurrence (67%) that was exactly the same as for the most basic test of deterioration (intact/non-intact).

Goal 5: To estimate the effect of dust lead measurement error on dust lead loadings.

In a subset of dwellings from all three sites in the study, side-by-side reliability samples were collected. Side-by-side dust samples in the home were used to estimate side-by-side variability for each sample type and site. All dust samples in the home (except additional side-by-side samples) were used to estimate between building variability and combined estimates of room/error variability for each sample type and site.

Using a combined estimate of room/error variability, observations were randomly generated from a log-normal distribution with these estimates of variability and various specified “true” average dust lead levels. This analysis was based on the assumption that there is some “true” unobservable dust lead level in a dwelling on a given surface type. Each dust sampling location was assumed to be equally representative of the true “unobservable” dust lead level in the dwelling on that surface type.

For the sample mean and maximum based on 1, 2 and 4 samples per dwelling, the following errors are evaluated:

(i) Type I (False Positive) Error = the probability that the <u>sample statistic fails</u> the dust lead standard given that the “<u>true” lead level is below the standard</u>.

(ii) Type II (False Negative) Error = the probability that the <u>sample statistic passes</u> the dust lead standard given that “<u>true” lead level is above the standard</u>.

The analyses generated Type I and Type II error estimates for each combination of site, surface type (uncarpeted and carpeted floors, window sills and window troughs), number of samples (1-5), and a prescribed set of dust lead standards. To simplify the presentation of these numerous results, a limited number of estimates are presented in Table ES-6. Estimates that represent significance levels 0.05, 0.10 and 0.20 are presented for each of the sites for floors and window sills. For comparative purposes, the effects of having only one or two samples collected in the dwelling are presented for floors in Milwaukee.

Results:

Table ES-6: Estimates of Upper and Lower Uncertainty Levels by Study Site, Surface Type, and Number of Samples

City	Surface	Standard	# of Samples	Confidence Level					
				95%		90%		80%	
				Lower	Upper	Lower	Upper	Lower	Upper
Milwaukee	Floor[1]	40	1	17	145	21	120	28	85
			2	20	99	23	83	30	67
			4	24	73	27	65	32	53
Balt. Co	Floor[1]	40	4	28	58	30	55	34	50
Milwaukee	Floor[1]	40	4	24	73	27	65	32	58
New York	Floor[1]	40	4	25	68	28	60	32	53
Balt. Co	Sill	250	4	120	630	150	540	190	440
Milwaukee	Sill	250	4	80	1630	115	1250	180	840
New York	Sill	250	4	105	940	140	740	175	560

[1] Central dust sampling location, carpets and bare floors combined

From the perspective of being most protective of a child's health, the upper uncertainty bounds in the table are of most interest. For example, the 80% upper uncertainty bound for a window sill dust sample of four rooms in New York City (see bottom row of table) was 560 $\mu g/ft^2$ when the window sill standard was 250 $\mu g/ft^2$. In practical terms, this means that if the "true" average lead level is 560 $\mu g/ft^2$, then there is a 20% chance that the sample mean will be below the standard of 250 $\mu g/ft^2$. These estimates are based on the good recovery rates achieved by the labs in this study. If the recovery rate is low, the variability effects could be compounded.

Sample variability may be just as harmful to the interests of a property owner and the affordability of housing. Using the sampling characteristics in the example above with the 80% lower uncertainty bound, approximately 20% of the time a home with a "true" lead level of 175 $\mu g/ft^2$ would fail the standard of 250 $\mu g/ft^2$. In other words, 20% of the time a dwelling with a true level 30% lower than the standard will fail the standard due to variability.

- **The results suggest that with additional samples in a dwelling, errors are less likely, but even with four samples the rate of error can be high.** For example, if 40 $\mu g/ft^2$ was established as a "health-based" standard for floors, these results suggest that that it may be appropriate to set an "action-level" below that standard to take into account the variability and be truly health protective.
- The high levels of variability for window sills (and window troughs) may help explain why these components were not predictive of blood lead outcomes. **With the level of variability, any sampling plan including window sill samples may have problems predicting risk.**

1.0 INTRODUCTION

In 1997, the U.S. Department of Housing and Urban Development (HUD) awarded the National Center for Healthy Housing ((NCHH), formerly the National Center for Lead-Safe Housing) a grant to assess HUD's risk assessment and lead hazard screen protocols found in the 1995 HUD *Guidelines for the Evaluation and Control of Lead-Based Paint in Housing.* Two hundred fifty-four dwellings with young children were enrolled in the study in three sites: Baltimore County, MD; Milwaukee, WI; and New York, NY. From June to October 1998, certified risk assessors conducted comprehensive risk assessments/paint inspections in these homes, and children's blood tests were matched with the results.

The primary goal of this study was to assess under what conditions HUD's risk assessment and lead hazard screening protocols are accurate predictors of children's lead exposure. The study attempted to identify ways to improve the accuracy of the protocols. To address these goals, NCHH conducted a detailed, multi-media environmental assessment of residential lead in a variety of housing and linked those results to children's blood lead levels. The resulting data set served as a test bed for a number of statistical analyses that address many of the key issues regarding the identification of housing that contributes to childhood lead poisoning.

The study had the following goals:

1. To assess the ability of the current HUD risk assessment protocols to predict dwelling units that are likely to house children having elevated blood lead levels, and assess the effect of modifying the protocols.
2. To assess the ability of the current HUD lead hazard screening protocols to predict the need for risk assessments, to predict dwelling units that are likely to house children having elevated blood lead levels and to assess the effect of modifying the protocols.
3. To describe the contribution of friction and impact surfaces to floor dust lead loadings.
4. To assess the ability of the current HUD paint film quality classification system to predict rooms and dwellings that are likely to have elevated dust lead loadings.
5. To estimate the effect of dust lead measurement error on dust lead loadings.

1.1 Background

Under the authority of Title X of the Housing and Community Development Act of 1992, HUD, in consultation with other Federal agencies, issued *Guidelines for the Evaluation and Control of Lead-Based Paint in Housing* [1]. The *Guidelines* were written to provide detailed, comprehensive technical information on how to identify lead-based paint hazards in housing and how to control such hazards safely and efficiently. One chapter (Chapter 5) was dedicated to the lead hazard risk assessment, a method of evaluation defined by Title X as "an on-site investigation to determine and report the existence, nature, severity and location of lead-based paint hazards in residential dwellings."

The authors of the *Guidelines* used the best scientific data available at the time and practical experience derived by insurers of Public Housing Authorities to develop the risk assessment protocols [2]. The *Guidelines* state that "These protocols represent the minimum recommended procedures for conducting risk assessments, and attempt to strike a balance between the need to have enough data to made informed decisions and the need to contain costs." The authors also used the experts' experience to develop a lower cost lead hazard screen risk assessment to serve as a "negative screen" for dwellings in good condition. Among professionals, there was a strong consensus that the protocols in the *Guidelines* represented the best expert judgment at the time, but there was also recognition that further research to validate the protocols was necessary. In July 1995, one month after the *Guidelines* were published, the Federal Task Force on Lead-Based Paint Hazard Reduction and Financing released a report that included research into the utility of the lead risk assessment and screening protocols recommended in the Guidelines as a key topic for investigation.

On November 27, 1996, the HUD Office of Lead Hazard Control issued a Notice of Funding Availability (NOFA) for Research to Improve the Evaluation and Control of Residential Lead-Based Paint Hazards. The NOFA recommended an investigation of the current risk assessment protocols as a strong candidate for research. After a competitive application process, NCHH was awarded a grant to research the topic.

In July 2000, a draft report on this study was submitted to HUD. After reviewing the report, HUD requested additional analyses and provided supplemental funding to complete these analyses in late 2001. This document takes into account regulations concerning lead-based paint risk assessment protocols and standards issued by HUD and the Environmental Protection Agency (EPA) in 1999 and 2001, respectively [3,4].

1.2 Overall Study Design and Limitations

The first and second goals of the risk assessment study, as recommended in the NOFA and adopted by NCHH, were to determine whether the *Guidelines'* risk assessment protocols and lead hazard screen protocols are predictive of children's lead exposure. These goals helped dictate the principal design elements for two of the three sites in the study. In Milwaukee and New York City, children's blood lead results, as reported by local blood lead registries and participating clinics, were used to identify pre-1950 dwellings eligible for the study. To obtain the target study population at these two study sites, a case-control methodology was used in which: half of the dwellings enrolled housed an EBL child (≥10 µg/dL) and the other half housed a child with a non-elevated blood lead level. The association of results of risk assessments conducted in these dwellings and the EBL status (EBL/Non-EBL) of the resident child in the dwelling was explored.

While NCHH's researchers believed that the case-control design was optimal to achieve the first two goals of study, they recognized that this design could not be used to assess the utility of all of the objectives of a risk assessment. While a principal objective of a risk assessment is to identify lead-based paint hazards that have the potential to poison a child, an equally important objective of a risk assessment is the final report that suggests appropriate hazard control options. The case-control study design implemented for this project was not intended to assess the latter objective. In fact, it is theoretically possible that a limited number of samples could identify a "high-risk" dwelling, but those limited samples might be insufficient for a risk assessor to accurately assess the lead hazard control needs of an owner. Additional studies will be required to determine how well the risk assessment protocols help an assessor make lead hazard control recommendations and how well risk assessors actually make these recommendations in the field.

A second limitation of the case-control design is that the outcome measure, blood lead level above/below 10 µg/dL, was determined at the start of the study. Because the study population was determined by this stratification, it would be difficult to define a different outcome measure at this time and still maintain a robust study. The ability to redefine high and low risk at 5 µg/dL or 15 µg/dL (or any other level) and maintain viable analyses is not possible.

Early in the study design process, it became apparent that the third site in the study, Baltimore County, Maryland did not have an accessible blood lead registry nor a blood lead screening rate sufficient to identify enough children for a four to six- month case-control study. NCHH elected to use a cross-sectional study design at this site. Potential study subjects were identified by the County based on a match of birth records and age of housing data from the tax assessor's office. Eligible families in post-1950 housing were invited by letter to participate in the study. Blood lead samples were drawn concurrently with the environmental sampling, so the population could not be selected on the basis of blood lead level. The original study design called for incorporating the Baltimore County data with the other sites in a final analysis based on the case-control design, but analytical concerns (described in Section 5.2.1) did not support merging the data. The Baltimore County data were therefore principally used in the analyses of the final three study goals: 1) to assess the contribution of friction and impact surfaces on floor dust lead levels, 2) to assess the current HUD paint film quality classification system, and 3) to estimate the effect of dust lead measurement error on dust lead loadings.

The focus of this study was on the HUD risk assessment and lead hazard screen protocols as defined by Chapter 5 of the *Guidelines*. Since it's publication date in 1995, however, the *Guidelines* have been clarified and updated through additional Federal publications and regulations. Late in 1995, the Environmental Protection Agency (EPA) released the *Lead-Based Paint Risk Assessment Model Curriculum* [5]. The *EPA Model Curriculum* is used in the training of all certified risk assessors. Because only certified risk assessors are allowed to officially conduct a HUD risk assessment, the *EPA Model Curriculum* was used as a clarifying supplement when the protocols in the *HUD Guidelines* appeared ambiguous. In 2001, EPA released regulations that changed the numeric standards for dust lead hazards and soil lead hazards. This risk assessment study presents the effectiveness of the risk assessment as measured by the new regulatory levels as well as by the original standards.

The study was designed to test the validity of the HUD risk assessment protocols and not a combination of a risk assessment and a paint inspection. When designing the study, however, NCHH determined that it would be most cost effective and least intrusive to include comprehensive XRF testing as part of the data collection process instead of relying on paint chip collection on deteriorated surfaces. To test the validity of the risk assessment protocols, the paint lead data were only used when deteriorated paint was present. In other analyses, such as the investigation of the effect of window friction on floor dust lead, paint lead levels of all surfaces, whether deteriorated or not, were included.

2.0 STUDY DESIGN

The Risk Assessment study collected data from Baltimore County, MD; Milwaukee; and New York City. These three jurisdictions were chosen because of the different ages and types of housing stock, and the presence of strong, capable local partners to help manage the study. The study population consisted of units housing a child, one to three years of age, who lived at the residence for at least six months prior to enrollment. A comprehensive set of environmental tests were taken in each home, including a visual inspection, XRF inspection, dust wipes, paint chips, soil and water samples. In addition, blood lead levels were collected or reported from one eligible child in the family, and a family interview was administered. These tests occurred within three weeks of each other, and all the data were collected within one five-month summer "season" in order to reduce confounding factors.

2.1 Enrollment Process Overview

The following enrollment procedures applied to all three jurisdictions. A unit was eligible for inclusion in this study if:

1) a child between 12 and 36 months of age had been in residence in their present home for at least six months prior to the study start date; and

2) that child had not been chelated.

3) that child did not have chronic health problems such as sickle cell anemia or chronic medical developmental problems

Only one eligible child per household was included in the study. Each jurisdiction had other conditions for eligibility (described in Appendix B). All families who participated in the study received a small monetary incentive that served to reimburse them for their time and any inconvenience related to study participation. Formal signed consent forms were obtained before any interviews or environmental tests took place. Finally, the environmental testing was completed in the home either prior to or concurrent with the family receiving information on the benefits of lead-specific cleaning, to reduce the likelihood of cleaning prior to the environmental testing. The original enrollment plan targeted 75 units in both New York City and Milwaukee, and 100 units built between 1950 and 1978 in Baltimore County. (More units were targeted in Baltimore County because the housing stock is less homogeneous.)

In New York City and Milwaukee, children and properties were recruited using the existing blood lead surveillance data bases or clinical records. Half the enrolled children had blood lead levels below 10 µg/dL, and half had blood lead levels equal to or greater than 10 µg/dL. This case-control study design facilitates the investigation of rare events - elevated blood lead status in the general population - and allows efficient use of resources to find differences between children who have elevated blood lead versus those who do not. Because blood lead testing in Baltimore County has been minimal, particularly for children living in post-1950 properties, a cross-sectional study design was used at this site. Specific enrollment procedures for each jurisdiction are described in Appendix B.

2.2 Environmental Data Collection Overview

All dwelling units in the study were tested in accordance with a standard environmental testing protocol described in the *Protocols for the Study of the HUD Risk Assessment and Lead Hazard Screen Procedures and their Effectiveness* (Appendix B). The environmental testing included a visual inspection, XRF inspection, dust wipe sampling, soil sampling, and water sampling. The environmental testing protocols were based on the risk assessment protocols described in Chapter 5 of the *Guidelines* and supplemented with additional data collection as summarized below.

2.2.1 Visual Inspection

The visual inspection included the standard review of building component conditions that is recorded on Form 5.1 of the *Guidelines*. An additional review of the common space condition was added to the form as well as a question about the general condition of interior floors.

To assess the influence of friction/impact surfaces systematically, the study went beyond the protocols found in the Guidelines, which states that "operating three or four windows, and three or four doors is usually adequate; it is not necessary to operate all windows and doors in the dwelling." The risk assessors documented all windows and doors at the dwelling and recorded the presence of paint, the condition of paint, whether the paint was rubbing or binding on each of these components. For windows, the risk assessors also recorded their general accessibility both in terms of potential dust blow-in (e.g., whether the window could be opened) and access of the child to the sill or trough (e.g., height off floor and presence of any barriers in front of window).

The risk assessment protocols also called for an assessment of the paint condition of all painted components in the dwelling. Since the study protocols required an XRF inspection, the visual assessment of paint was incorporated into the paint inspection.

2.2.2 Paint Inspection

The risk assessment protocols in the *Guidelines* call for the risk assessor to determine whether any deteriorated paint is lead-based. The paint may be tested by XRF instrument or by paint chip analysis. After considering all of the goals of the study, NCHH determined that intact painted surfaces would also have to be tested to assess the HUD paint classification system and the influence of friction/impact surfaces.

The HUD paint inspection protocols as revised in 1997 were used as the basis of the paint inspection. However, five modifications were made to the paint inspection protocols to better achieve the objectives of the study and gain some cost and time efficiencies:

- Only one reading was taken from each testing combination, *including walls.*
- For every window tested, five window components were tested (if accessible): window sash, window jamb, window casing/apron, window sill and window trough.

- Trim moldings (baseboards, chair rails, crown moldings) were treated as one testing combination.
- Painted floors were tested in each room.
- Only components that could not be tested with the XRF instrument AND had loose, peeling paint were tested by paint chip analysis.

The last exception listed was implemented after a local Institutional Review Board objected to destruction of intact paint. Because the protocols only allowed XRF instruments that had no inconclusive ranges, the impact of this exception was minimal. Only one unit required paint chip testing.

2.2.3 Dust Sampling

The study dust sampling procedures were developed in accordance with the dust sampling protocols for composite and single-surface dust wipe samples found in the HUD *Guidelines*. Although the *Guidelines* allow risk assessors to use their professional judgment when selecting sampling locations, the study sampling plan designated precise sampling locations to improve the comparability of results across study locations. However, the risk assessors were required to note which of the designated locations they would have sampled had they been allowed to use their professional judgment.

In each dwelling, single surface dust wipe samples were collected in each of the following rooms:

- living room
- kitchen
- bathroom
- index child's bedroom
- second child's bedroom, and
- index child's play room (if not already sampled).

In each of these rooms, one sample was collected in each of the following locations:

- window sill
- window trough
- entryway floor
- perimeter floor
- window floor
- central floor

An additional sample was collected from the entry to the dwelling unit, and up to four common area dust samples were collected in multifamily common spaces.

In each dwelling, composite dust wipe samples were collected from up to six locations:

- uncarpeted floor – central
- uncarpeted floor – under window
- carpeted floor – central
- carpeted floor – under window
- window sill, and
- window trough

Each composite sample ideally contained four sub-samples and no less than two samples. The number of subsamples depended on the number of appropriate sampling locations available. The exact room locations for the composite samples are described in detail in the study protocols.

2.2.4 Reliability Dust Sampling

To meet the fifth goal of assessing the impact of measurement error on dust sample results, additional side-by-side sampling was conducted in a subset of dwellings. Fifty-nine dwellings in Milwaukee and forty-nine dwellings in New York City were selected for reliability dust sampling. This additional sampling consisted of:

- Dividing one window sill (not sampled as part of the standard dust lead sampling) into three sections (left, center, right) and sampling each section separately.
- Dividing one window trough (not sampled as part of the standard dust lead sampling) into three sections (left, center, right) and sampling each section separately.
- Taking two additional *carpeted* central floor samples on either side of a standard *carpeted* central floor sample.
- Taking two additional *uncarpeted* central floor samples on either side of a standard *uncarpeted* central floor sample.

2.2.5 Soil & Water Sampling

The study soil protocols followed the *Guidelines* protocols for the collection of a play area and perimeter composite soil samples. An additional composite soil sample was collected from the streetside curb area of the property.

The study water protocols followed the EPA water sampling protocols (as directed in the *HUD Guidelines*) without modification. Although water sampling is not included in the recommended risk assessment protocols described in the *Guidelines*, NCHH included it in this study since water is a possible residential exposure source.

2.3 Blood Lead Overview:

As described in Appendix B, blood lead collection in Milwaukee and New York City was done outside of the direct control of NCHH or its consultants. The labs that performed the blood lead analyses were monitored under the QA/QC plan for the study. In New York City, the blood lead collection methodology was restricted to venous samples. In Milwaukee, the study originally planned to restrict the eligible samples to venous samples, but enrollment proved difficult using this criteria. Therefore, an investigation of routine blood screening results from capillary (fingerstick) sampling and confirmatory venous sampling in Milwaukee was undertaken to determine whether capillary results could be used. The results were determined to be very consistent, although the capillary results were almost 6 percent higher than the venous results. To account for the bias, enrollment criteria for EBL children and non-EBL children were set at ≥ 13 μg/dl and ≤ 9 μg/dl, respectively. Twenty-two percent (17 of 79) of the blood lead samples collected in Milwaukee were collected by a capillary draw.

In Baltimore County, NCHH maintained direct control over the blood samples collected. Blood lead specimens were collected in the home by a phlebotomist under contract with NCHH. The nurse-phlebotomist was directed to attempt to collect a venous sample, but if unable, the blood could be drawn using capillary procedures. Eighteen of the 100 samples collected in Baltimore County were collected by a capillary draw. Specific details about the collection, storage and shipping of the blood lead samples can be found in the study protocols.

2.4 Household Interview Overview

A Household Questionnaire was completed with a resident in each enrolled dwelling unit. The primary reason for the household questionnaire was to better understand other non-building related sources of lead, including activities and hobbies performed by occupants at the home, remodeling and renovation activities in or near the home, occupant occupations, use of ceramics and home remedies and water usage. Room-specific information was also collected on cleaning practices (including a visual observation) and availability of cleaning equipment. Other demographic information included household composition and family income. This information helped to identify and quantify variables that could modify or confound the blood lead results.

The questionnaire was also intended to elicit specific information about the enrolled child and the child's habits to help identify factors that may affect the child's blood lead levels, including the length of time spent inside the home, time spent in the homes of others including day care, and time spent out-of-doors; behavioral patterns including mouthing, child activity at an identified window, and achievement of child developmental milestones; and child nutrition including an assessment of milk and fluid intake and use of vitamins.

The interview was conducted in person at the home where the family resides, following accepted practices of formal interviewing. The program staff interviewed an adult respondent who was a principal caregiver for the child who was enrolled. If needed, additional follow-up was done with the family by telephone to clarify responses to the interview. The Household Questionnaire was available in English and Spanish. For two Cambodian participants who spoke neither English nor Spanish, household information was obtained using experienced interviewers fluent in the participant’s language, Hmong.

2.5 Quality Assurance/Quality Control (QA/QC) Overview

2.5.1 Field Quality Control

All field personnel participating in the study were required to meet certification requirements (e.g., all risk assessor's were required to be EPA certified and have experience). The field personnel were trained on the study procedures by NCHH. During the data collection period, each of the personnel was observed in the field by NCHH staff or consultants at least three times. Reports were prepared assessing performance and compliance with the study protocols. Field personnel were notified how to improve compliance. Overall performance was satisfactory.

A pair of risk assessors conducted all risk assessments. In Milwaukee, a risk assessor identified two instances of gross non-compliance with the study protocols by his partner and notified NCHH. NCHH immediately removed the problem risk assessor from the project. The data in question was removed from the study dataset. The integrity of other data collected by the risk assessor was determined to be uncompromised.

2.5.2 Lab Quality

The University of Cincinnati Hematology and Environmental Laboratory was hired to serve as the QA/QC officer for sampling issues on this project. A comprehensive QA/QC plan was developed at the beginning of the project. A central laboratory was selected to conduct all environmental sample analyses. On a regular basis, blind blank composite and single-surface dust wipe samples and blind spiked samples for dust, soil and water samples were submitted in the shipments of field samples to the central laboratory. When results of the spiked samples fell outside of the QA/QC error limits established in the QA/QC plan, the QA/QC officer worked with the central lab to identify the problem. In some cases, the lab was requested to reanalyze a batch of samples. The overall performance of the laboratory was satisfactory.

The QA/QC officer was also responsible for submitting on a monthly basis two blind field control samples to the laboratories analyzing blood specimens in Milwaukee and New York City. In Baltimore County, the local Program Manager submitted the blind field control samples on a regular basis with the other samples. As with the environmental lab, the performance of each of the laboratories analyzing blood was considered satisfactory.

Detailed charts documenting laboratory performance are presented in Appendix E.

2.5.3 Data Quality

At each of the three sites, a local Program Manager was assigned the task of reviewing all field collection forms. Correctable data reporting errors were returned to the risk assessor for correction. Uncorrectable protocol errors were noted and then reviewed with the appropriate field personnel to prevent reoccurrence. After corrections were made, the local Program Manager shipped forms to NCHH for final review and data entry.

At NCHH, the developer of the study protocols reviewed approximately one-third of the field collection forms. Local Program Managers were notified when additional errors were identified. When possible, those errors were corrected at that time. The forms were then double data entered in Jetform's FormFlow data entry system [6]. Computer logic checks were run as part of the entry process and after the data were entered. Invalid data were corrected or deleted. Less than ten data points had to be removed because of invalid responses. All statistical analyses were performed using SAS/STAT® software [7].

3.0 SUMMARY OF THE FIELD STUDY

Pilot testing of the study protocols took place in March and April 1998. Local Program Managers and risk assessors were then trained in the final study procedures in May, with enrollment and testing beginning at the end of May. Final testing concluded at the end of October 1998. Program Managers enrolled 254 dwelling units where risk assessments were performed. Table 3.1 lists the distribution of enrolled units.

Table 3.1: Number of Dwellings Enrolled by Site and Elevated Blood Status

Site	Non-EBL Child Present (< 10 μg/dL)	EBL Child Present (≥ 10 μg/dL)	All Dwellings
Baltimore County	99	1	100
Milwaukee	42	37	80[1]
New York City	35	39	74
Total	176	77	254

[1] One unit was enrolled and tested based on the verbal report of a blood lead test, but the blood lead result was never confirmed by electronic report.

In Milwaukee and New York City, the recruitment goal of the case-control study design was achieved with 153 enrolled dwellings housing approximately equal numbers of non-EBL children (<10 μg/dL) and EBL children (≥10 μg/dL) in each site. After field data collection was complete and the data were entered, the study identified 17 dwellings where either the index child's age (11 dwellings) or the time between blood and dust sampling (6 dwellings) fell outside of the study's eligibility criteria. The study considered the results and determined that the criteria could be adjusted without harm to the integrity of the study. The child's age restrictions were revised from between 12-36 months to between 10-38 months of age, inclusive, and time of blood sampling relative to dust sampling was revised from +/- 21 days to 28 days before to 21 days after. Four of the 17 dwellings were ineligible based on the revised criteria (see Section 5.1.4). All of the dwellings in Milwaukee and New York City were constructed prior to 1950 with the exception of two dwellings in Milwaukee. These two dwellings were also considered ineligible.

In Baltimore County, where the limited screening data precluded the use of a case-control design, the cross-sectional approach resulted in a sample dominated by dwellings that housed children with blood lead levels below 10 μg/dL. Since the study population was made up solely of dwellings that were built after 1950, and had a largely white population, the results are consistent with the CDC's National Health and Nutrition Examination survey (NHANES) for 1991-1994. The NHANES survey estimated that 1.4% of white children living in housing built between 1946 and 1973 would have a blood lead level above 10 μg/dL. [8]

4.0 INTRODUCTION TO STATISTICAL ANALYSIS RESULTS

The next three sections and Appendix A of the report describe the outcomes of the analyses that examined the ten objectives in the study's Statistical Analysis Plan. The analyses for these objectives build upon themselves, first examining the validity of the current risk assessment and lead hazard screen protocols as tools to identify residences likely to house EBL children; then exploring individual factors influencing risk assessment performance such as paint condition, dust sample location and dust sampling variability; and finally, exploring alternative risk assessment protocols.

Section 5: Performance of the Current Risk Assessment and Lead Hazard Screen

- Objective #1. Assess the ability of the current HUD risk assessment protocol to predict dwelling units that are likely to house children having elevated blood lead levels.
- Objective #2. Assess the ability of the lead hazard screen protocol to predict the need to conduct a risk assessment. Assess the ability of the HUD lead hazard screen protocol to predict dwelling units that are likely to house children having elevated blood lead levels.

Section 6: Exploration of Individual Factors Influencing Risk Assessment Performance

- Objective #3. Explore the relationship between single and composite dust lead loading standards
- Objective #4. Describe the contribution of friction/impact surfaces, blow-in and track-in to dust lead loading
- Objective #5. Assess the ability of the current HUD paint film classification system to predict rooms and dwellings that are likely to have elevated dust lead loading levels
- Objective #6. Explore the Components of Variation in Dust Lead Sampling
- Objective #7. Explore the effect of modifying dust sampling protocols in the prediction of dwelling units that are likely to house children having elevated blood lead levels.

Section 7: Investigation of Better Risk Assessment Tools[1]

- Objective #8. Explore the effect of modifying the HUD risk assessment protocols in the prediction of dwelling units that are likely to house children having elevated blood lead levels.

Appendix A: Questionnaire Results

- Objective #10: Identify demographic, family or child characteristics that affect the effectiveness of a risk assessment in predicting children's blood lead levels.

---

[1] The researchers dropped Objective #9 (Assess the effect of modifying the lead hazard screen protocols to predict the need for a risk assessment and; to predict dwelling units that are likely to house children having elevated blood lead levels) after it was determined that simple modifications would not improve the performance of the Screen.

The analysis plan called for the analyses to build upon themselves, but findings late in the analysis process resulted in the study deviating somewhat from the plan. As is described in detail in Section 7, the study team found that although the study design was very similar for Milwaukee and New York City (i.e., case-control study of children in pre-1950 dwellings), the relationships between environmental lead levels at the dwellings and the children's blood lead status (EBL/non-EBL) were very different for the two sites. Some lead sources in the home environment were significantly related to blood lead status in Milwaukee, but this finding was not true in New York City. Therefore, combining data from Milwaukee and New York City for analysis, as was done in the early stages of the analysis, proved to cloud our understanding of how lead in the home is related to elevated blood lead levels. Instead of using the early analyses of the relationships of dust lead and paint lead with blood lead to help design the final analysis exploring more optimal risk assessment protocols, some of the earlier analyses in Section 6 were rerun after there was strong evidence that data from Milwaukee and New York should be examined separately. The study team recognizes that had the analyses in Section 6 been completed first, some of the variables used in the Section 7 may have been different. While the analyses do not necessarily build on themselves, the findings throughout this report offer important new information on the relationship between lead in the home environment and children's blood lead levels.

The study team determined that it was appropriate to continue to present the findings in Section 5 with data from Milwaukee and New York City combined. By retaining the combined analyses in this Section, two important themes are presented that must be recognized by policy makers and researchers:

First, the dwelling was not necessarily the source of the child's lead exposure. The study design included residency requirements for the child (i.e., at least six months at dwelling) to increase the likelihood that the child's blood lead level was a function of the child's home environment, but this did not guarantee that the child received the majority of his or her lead exposure from that environment. This matches real life situations where a lead-based paint risk assessment may not be predictive of an individual child's risk of lead exposure if the child is exposed at a different location or from a non-residential source. Thus, findings such as those presented in Section 5 which identify no significant relationship between the risk assessment results and the enrolled child's blood lead status, are not necessarily an indictment of the validity of current protocols. Instead, these findings may suggest that while it may be possible to improve the predictive power of lead risk assessments, a lead risk assessment by itself should not be expected to identify all dwellings where an child with an elevated blood lead level may reside.

Second, the results of Section 5 offer insights on how results can be affected when findings from locations with very different lead exposure patterns are combined. While the findings in Section 7 suggest that with some adjustments the current risk assessment protocols appear to be a valid measure of the risks of lead in the home environment, a very different conclusion might be drawn from examining the combined data. Researchers exploring the relationships between environmental lead in the home and children's lead exposure must understand the lead exposure patterns for each site in their dataset before pursuing their final analyses.

5.0 PERFORMANCE OF THE CURRENT RISK ASSESSMENT AND LEAD HAZARD SCREEN (OBJECTIVES #1 AND #2)

5.1 Methodology

5.1.1 Type of Statistical Analysis

A standard method to assess the accuracy of a diagnostic test is to examine the performance characteristics of the test, using four probability measures: sensitivity, specificity, positive predictive value, and negative predictive value. For the purposes of this study, the four terms are defined as follows:

- *Sensitivity (or True Positive Rate)*: Probability that a dwelling unit fails an environmental assessment given that there is a resident child with an elevated blood lead concentration.
- *Specificity (or True Negative Rate)*: Probability that a dwelling unit passes an environmental assessment given that a resident child does not have an elevated blood lead concentration.
- *Positive Predictive Value*: Probability of a resident child having an elevated blood lead concentration given that the dwelling unit failed the environmental assessment.
- *Negative Predictive Value*: Probability of a resident child not having an elevated blood lead concentration given that the dwelling unit passed the environmental assessment.

In addition to the probability measures listed above, the analysis presents the confidence intervals for the performance characteristics and a statistical test of independence between the environmental assessment result (pass/fail) and the presence or absence of a child with an elevated blood lead level. The latter test is critical to the analysis because it indicates the relative strength of any diagnostic test. A result with a p-value less than 0.05 indicated that the environmental assessment and the blood lead status of a child were not independent from each other. The test assesses whether the probability of a dwelling without a child with an elevated blood lead failing the risk assessment is the same as the probability of a dwelling with an EBL child failing (i.e., an undesirable outcome).

5.1.2 Diagnostic Tests Undergoing Analysis

Although the HUD risk assessment sampling protocols have remained largely unchanged since the *Guidelines* were issued in 1995, the measures used to indicate lead hazards have been modified. Both HUD and EPA have issued final regulations that have changed the hazard definitions for dust lead hazard and soil lead hazards. The performance characteristics analysis presented in this Section assesses the predictive value of the HUD risk assessment protocols using the standards stated in the *Guidelines* as well as the final standards published by EPA since the *Guidelines* were issued. The analysis also presents the performance characteristics for the lead hazard screen using the 1995 standards and the 1999 HUD interim standards.

The environmental testing protocols that were considered are summarized in Table 5.1. The sources of the standards for the protocols are described below.

5.1.2.1 Risk Assessment Protocols

HUD Guidelines: On September 11, 1995, three months after the *HUD Guidelines* were issued, EPA published a guidance letter in the Federal Register entitled, *Guidance on Identification of Lead-Based Paint Hazards; Notice.* That document introduced into the public record a 1994 memorandum from Lynn R. Goldman, Assistant Administrator for Prevention, Pesticides and Toxic Substances, that established temporary standards for lead-based paint hazards. The EPA has rulemaking authority under Section 403 of Title IV of the Toxic Substances Control Act (TSCA) to promulgate regulations concerning the definitions of lead-based paint hazards. These protocol standards are considered here the 1995 **HUD Guidelines** standards.

The HUD *Guidelines* accepted the use of both single-surface and composite dust sampling when conducting risk assessments. EPA also supported this position in its final regulatory rules for the *Requirements for Lead Based Paint Activities in Target Housing and Child Occupied Facilities* (Federal Register, August 29, 1996). This rule delineated the activities approved for certified lead professionals. At the time the HUD *Guidelines* were written, both agencies agreed that when using single-surface sampling, any individual sample result that falls above the dust lead standard indicates a hazard. In effect, an at-risk home was identified (i.e., the home “failed” the risk assessment) based on the maximum dust lead level on any component type (floor, sill, or trough). When using composite sampling, the results effectively dictate that if the sample *average* falls above the dust lead standard, then a hazard is indicated. Assuming just one composite sample per component type in a home, a home “fails” the risk assessment based on the average dust lead level. The analysis presents performance characteristics for both outcome measures under **HUD Guidelines – Average** and **HUD Guidelines –Maximum.**

Note: In order to avoid the additional confounding factor of comparing composite results with single-surface results, the study uses the arithmetic mean of the single-surface samples in the analysis of HUD Guidelines – Average. Additionally, the 1994 EPA Guidance document only specified a dust standard for uncarpeted floors, while the HUD Guidelines allowed the standard of 100 $\mu g/ft^2$ to be used on carpeted floors as well as smooth uncarpeted floors. For these analyses, the floor standard was applied to all floors.

HUD Interim: On September 15, 1999, HUD issued its final rule on the *Requirements for Notification, Evaluation and Reduction of Lead-Based Paint Hazards in Federally Owned Residential Property and Housing Receiving Federal Assistance* (Federal Register). These rules were authorized under Sections 1012 and 1013 of the Residential Lead-Based Paint Hazard Reduction Act of 1992. The rule established interim standards for lead hazards to be used during the period between the effective date of the regulations (September 15, 2000) and the implementation of the final EPA Section 403 regulations. The HUD 1012/1013 regulations lowered the floor dust lead standard from 100 to 40 $\mu g/ft^2$ and the window sill dust lead standard

Table 5.1: Environmental Testing Protocols Studied

	Dust Lead Loading (μg/ft^2) Standard			**Poor[1] Paint Lead (mg/cm^2)**	**Bare Soil[4] Lead (PPM) Standard**		
Assessment	**Floor**	**Window Sill**	**Window Trough**		**Play Area**	**Yard Perimeter**	**Remainder of Yard**
Risk Assessment							
HUD Guidelines (Average dust)	100	500	800	1	400	2,000	NA
HUD Guidelines (Maximum dust)	100	500	800	1	400	2,000	NA
HUD Interim (Maximum dust)	40	250	NA	1	400	2,000	NA
EPA Current (Average dust)	40	250	NA	1	400	NA	1,200
Lead Hazard Screen[5,6]							
HUD Original	50[2]	NA	400	1	NA[3]	NA[3]	NA
EPA Current	50[2]	250	NA	1	NA	NA	NA
HUD Current	25[2]	125	NA	1	400	2000	NA

NA=Not applicable

[1] According to HUD's definition of "poor" condition.
[2] Uncarpeted floors only
[3] Although the *Guidelines* specify testing soil if exterior paint chips present, in this analysis, soil is excluded since only 4 units that qualify for the screen have paint chips present.
[4] In the study, soil covering was rated as: No bare (0-10%), Small amount bare (10-40%), half bare (40-60%), mostly bare (60-90%) or all bare (90-100%). Although this does not exactly correspond to federal guidance, in this analysis, the standard applies unless there is "no bare" in the specified exterior region.
[5] The screen should be applied only if conditions outlined in the HUD Guidelines are met. For this study, the criteria were that there are <2 system deteriorations (as defined by Form 5.1 of the Guidelines) and ≤5 interior painted testing combinations are in poor condition.
[6] Composite samples were used for the screen. For floors, the central uncarpeted location was used.

from 500 to 250 µg/ft^2. The rule dropped the testing of window troughs for risk assessment purposes. The regulations maintained the same method of determining dust lead hazards; an at-risk home was identified based on the maximum single-surface dust lead level or the mean composite dust lead level. The soil lead standards were not changed.

EPA Current: On January 5, 2001, the EPA issued the final rules for the *Identification of Dangerous Levels of Lead* as required under section 403 of Title IV of TSCA. These rules were generally in agreement with the HUD rules. EPA determined that dust lead hazards existed based on a floor standard of 40 µg/ft^2 and a window sill standard of 250 µg/ft^2. While HUD originally retained the use of the maximum dust lead level of a component to identify high-risk homes when single-surface wipe sampling is conducted, EPA regulations established the average dust lead level as a measure of risk.

EPA maintained the same soil standard for play areas, but revised the other soil hazard standard. Instead of determining the existence of a hazard based on a perimeter sample, EPA broadened the sampling area to the remainder of the yard, including the perimeter. EPA set the soil standard for the remainder of the yard at 1,200 ppm.

5.1.2.2 Lead Hazard Screen Risk Assessments

HUD Original Screen: The HUD *Guidelines* contain protocols for an abbreviated environmental assessment that would act as a negative screen for risk assessors. The Lead Hazard Screen Risk Assessment is intended to be used in dwellings in good condition that are likely to pass a full risk assessment. By investing in a screen, property owners can determine that their dwellings are low-risk without incurring the expense of a complete risk assessment. EPA recognized the lead hazard screen protocols in its final Section 402/404 rule.

The lead hazard screen protocols in the HUD *Guidelines* call for a visual assessment of paint hazards and an analysis of two composite wipe samples taken from floors and window troughs. The dust lead standard for the screen is set at half the standards set for the risk assessment. HUD offers no opinion as to whether the floor composite sample should be collected from carpeted or uncarpeted floors, but does not allow these surface types to be mixed in one composite sample. The HUD protocols state that soil sampling is optional, but should be done if paint chips are observed on the soil. (Because paint chips were discovered at only 4 of 62 eligible study dwellings (6%), soil sampling is not included in the analysis of these protocols.)

EPA Original Screen: When EPA issued its certification rules in 1996, its protocols deviated slightly from the HUD Guidelines. While both agencies agree that the lead hazard screen consists of a visual assessment of paint hazards and an analysis of two composite wipe samples taken from floors and a window component, EPA leaves the decision about the window location (i.e., sill vs. trough) to the risk assessor's professional judgment. Because EPA only set standards for uncarpeted floors, floor samples are required to be from uncarpeted locations. EPA has no requirements for soil sampling.

HUD Current Screen: HUD's final rules for Section 1012/1013 changed its original lead hazard screen protocols in four areas. First, it allows the use of either composite sampling or the average of single-surface sampling of dust. Second, it requires window sill dust lead samples instead of window trough samples. Third, it changes the dust lead standards to 25 $\mu g/ft^2$ for floors and 125 $\mu g/ft^2$ for window sills. These standards are essentially in-line with using a standard of half the risk assessment standard, but uses 25 $\mu g/ft^2$ instead of 20 $\mu g/ft^2$ on floors out of concern for practical laboratory detection limits. Finally, HUD final rules make soil sampling of play areas and perimeters of buildings mandatory instead of optional.

Although the Section 1012/1013 rules allow single-surface samples to be used, the performance characteristics analyses for all three lead hazard screen protocols were run using only composite samples. For each screen protocol, the floor sample was represented by the uncarpeted composite floor sample that was collected from the center of each room.

For all three lead hazard screen protocols, the issuing agency strongly recommends that the lead hazard screen be conducted at dwellings that have a good chance of passing the test. The purpose of the screen is to offer the property owner a lead evaluation tool to identify low-risk dwellings at less cost than the complete risk assessment. Since the screen protocols require the risk assessor to return to the property for further testing and possible development of a lead hazard control plan if the screen fails, it is not be cost-effective to conduct lead hazard screens in dwellings that are likely to fail. The HUD *Guidelines* suggest that a dwelling is likely to fail the screen if the dwelling is in poor condition. EPA suggests that a dwelling is likely to fail the screen if the dwelling was built prior to 1960. HUD now recognizes both age and condition as indicators of likely failure and recommends the screen only for post-1959 dwellings in good condition.

For reasons to be discussed in Section 5.2.1, it was analytically inappropriate to include Baltimore County's post-1959 dwellings in the performance characteristics analyses. In order to assess the performance characteristics of the lead hazard screen, the analyses could not exclude dwellings on the basis of age and only restricted eligibility by the building condition. Buildings in poor condition were not considered. Because there is some ambiguity about the definition of a building in poor condition, the study defined a dwelling as being in poor condition if:

1. More than one building component on *Guidelines Form 5.1* was deteriorated[2].
2. More than five painted testing combinations were in poor condition.

The first criterion was based on the footnote to Form 5.1 in the *Guidelines*. The second criterion is alluded to in the *Guidelines* (p. 5-25) and is specifically referenced in EPA *Lead-Based Paint Risk Assessment Model Curriculum* (p. 9-6).

5.1.3 Additional Analytical Decisions Concerning the Use of Environmental Samples

---

[2] The Form 5.1 in Appendix 8.2 of the HUD Guidelines states that a dwelling is in poor condition if *four* or more building components are deteriorated, but this study used the guidance found on the Form 5.1s in Chapter 5 and in Appendix 8.1 that define poor as *two* or more deteriorations.

Beyond the specific protocol-based decisions discussed above, the study made a number of decisions concerning the use of environmental samples in the analyses. These decisions are described below by sample type: dust, paint, and soil.

5.1.3.1 Dust

All risk assessment protocols allowed the risk assessors to use their own professional judgment to select the exact rooms and locations of dust sampling. For example, the *Guidelines* state:

> These *Guidelines* provide advice on deciding which rooms to sample and which components to sample within rooms. However, only general guidance can be offered on exactly *where* samples should be collected. The exact spot to be sampled should be chosen based on the risk assessor's visual observations and the results of any resident interviews and use patterns. (p. 5-15)

This study, however, required risk assessors to collect up to 37 single-surface dust samples within a dwelling unit (as well as up to six common area samples) from specific locations (see Section 2.2.3). The risk assessors were then asked to identify 6-8 samples in the field that they would have selected as part of a risk assessment had this not been a study. These 6-8 sample locations, which are described in Section 5.2.2.2, were used to test the risk assessment study protocols.

The ratio of floors-to-sills-to-troughs assessed was purely at the discretion of the risk assessor. Based on his/her professional judgment, the risk assessor may have determined that a component type could not or should not be sampled. For these analyses, if a dust sample component type was not one of the identified risk assessment sampling locations, it was assumed to have passed. For example, if no window sill was selected as a risk assessment sampling site, the dwelling was assumed to "pass" for sills.

5.1.3.2 Paint

A standard risk assessment requires the testing of only deteriorated lead-based paint. While this study included complete XRF testing, only painted surfaces classified as deteriorated and leaded (1 mg/cm^2 or greater) were included in the analysis of risk assessment performance characteristics.

The exact definition of a deteriorated painted surface has been a topic of much discussion since Title X was passed. For this study, the *Guidelines* definition of "poor" paint condition was used as the definition of a lead-based paint hazard that would trigger the "failure" of a risk assessment. Paint is in "poor" condition if there is deterioration (chipping, peeling, flaking, etc.) on:

a) more than ten square feet of large exterior surfaces,

b) more than two square feet of large interior surfaces, or
c) more than 10 percent of the total surface area of small interior surfaces.

The *Guidelines* explicitly state that only lead-based paint in "poor" condition is a lead-based paint hazard. Based on this policy position, the study used "poor" as the paint condition standard. It must be noted, however, that in the final Section 1012/1013 rules issued in September 1999, HUD removed the *de minimus* standard for deteriorated lead-based paint. EPA also removed the *de minimus* standard in its final Section 403 rules. HUD and EPA now consider that all non-intact leaded paint is deteriorated and is a lead-based paint hazard. While "poor" condition paint is retained as the definition of a risk assessment failure in the analysis of the **HUD Interim** and **EPA Current** protocols, the ramifications of switching to the new definition of deteriorated paint will be discussed. Furthermore, the predictive power of the complete HUD paint lead classification system will be discussed in further detail in Section 6.3.

5.1.3.3 Soil

Title X stated that lead-in-soil is to be considered a hazard when the soil is "bare." The exact definition of bare soil has been under discussion since the law was passed. Because even well-maintained yards often have patches of soil where the ground cover does not grow, the regulatory agencies have varied in their definitions of "bare" soil locations.

The HUD *Guidelines* state that, with the exception of play areas, sampled areas have to contain more than nine square feet of bare leaded soil to be considered a lead-based paint hazard. The same standard has been retained in the final Section 1012/1013 rules. HUD considers any bare leaded soil in a play area a hazard. In the EPA Section 403 rules, the agency decided not to set a *de minimus* standard for bare soil in any part of the property.

The risk assessment study protocols did not call for the risk assessor to provide an absolute measure of the amount of bare soil. Instead, the risk assessor classified the soil covering as: No bare (0-10%), Small amount bare (10-40%), Half bare (40-60%), Mostly bare (60-90%), or All bare (90-100%). For these analyses, the decision was made to exclude the No Bare category from consideration of any soil hazards. This decision may yield slight over counts of the number of non-play area surfaces that HUD would consider as hazards, and slight undercounts of the number of play areas (for HUD) and general yard areas (for EPA) that the agencies would consider as hazards. However, the analyses in the study will demonstrate that the study's definition of bare soil has a limited impact on the predictive performance of the current risk assessment protocols.

5.1.4 Eligible Dwellings for Analysis

As described in Section 3, 254 dwelling units were enrolled in the study: 100 in Baltimore County, 80 in Milwaukee and 74 in New York City. Of these units, seven units (all from Milwaukee) were excluded from the performance characteristics analyses because they did not meet the enrollment criteria.[3]

Although all of the remaining 247 dwellings had dust, paint and soil assessed, risk assessors failed to identify at least six dust sampling locations that they would have sampled at 17 dwellings: 3 in Baltimore County, 10 in Milwaukee and 4 in New York City. Because identification of these locations was necessary to perform the performance characteristics analyses as defined for the study, all 17 dwellings were excluded from these analyses. In addition, risk assessors in New York City failed to assess one yard-play area (i.e., no soil was collected, nor was it identified as inaccessible). A total of 229 dwellings were deemed eligible for the analysis of risk assessment performance characteristics: 97 in Baltimore County, 63 in Milwaukee and 69 in New York City.

Table 5.2: Identification of Dwelling Units Eligible for Analysis

Criteria	Baltimore County	Milwaukee	New York City	Total
Number of Units Enrolled/Tested	100	80	74	254
Number of Units Meeting Basic Enrollment Criteria	100	73[a]	74	247
Number of Units Eligible for Risk Assessment Analysis	97	63	70 (69[c])	230[b] (229[c])
Number of Units Eligible for Lead Hazard Screen Analysis	58	15	38	111[d]

a) Seven units excluded because enrollment criteria were not met
b) Seventeen units excluded because risk assessors failed to identify at least 6 dust sampling locations
c) One unit excluded from analyses with soil because risk assessors failed to assess play area soil
d) All units in poor condition were excluded from lead hazard screen analysis

---

[3] One dwelling lacked an associated blood lead result, four units had blood drawn more than four weeks before or three weeks after the environmental testing, and two units in Milwaukee were built after 1950.

In the analyses of the lead hazard screen protocols, composite dust samples were used, so the risk assessor's designation of sampling locations was not required. Instead, the main cause for data exclusion was the criterion that dwellings be in good condition. One hundred thirty dwellings were excluded because they were classified in poor condition. Three additional dwellings from New York City were excluded because composite window trough dust results were missing, while three dwellings from Baltimore County were excluded because of missing composite window sill results. As shown in Table 5.2, 111 dwellings eligible for use in the analyses of the lead hazard screen protocols.

5.2 Findings

5.2.1 Preliminary Considerations

Under the original analysis plan, the performance characteristics were to be considered for all eligible dwellings combined. Before conducting the analyses in this manner, a test of independence was performed on the combined data and separately, by site. The results were surprising and troubling. When all data were combined, the test demonstrated (as one would hope) that the results of the HUD Guidelines protocol with average dust and the blood lead outcomes (EBL/no EBL) are marginally related (p-value= 0.08). In other words, the test indicated that there is a statistical relationship between a risk assessment and the presence of a lead-poisoned child at p=0.08. However, when the data were analyzed by site, the risk assessment results and the blood lead outcomes were independent (p-value=0.443, 0.492 and 0.452 for Baltimore, Milwaukee and New York City, respectively). This set of findings, which in statistics is called the Simpson's Paradox, indicated that the study results would be flawed if the data were combined [9].

The site-by-site analyses indicated that in Baltimore County, where one child in 100 had an elevated blood lead level, approximately half of the homes failed the risk assessment (48-57% depending on the protocol). In Milwaukee and New York where by design about half of the children in the study had an elevated blood lead level, at least two thirds of the dwellings failed the risk assessment (67-97%). Although no individual site results could demonstrate that the risk assessment protocols have a statistically significant relationship with blood lead outcome, the combined dataset appeared to indicate a statistically significant relationship.

When considering alternatives to the combined analysis, feasibility issues precluded a separate performance characteristics analysis or any other type of analysis of the Baltimore County data. With only one child having an elevated blood lead level, there were not enough outcomes in each cell (EBL/no EBL) to produce reasonable results. For the Milwaukee and New York City data, it was determined that the data could be combined since the study design was the same (case-control study in pre-1950 buildings) and the outcomes were similar. Therefore, in this section of the report, the findings for Baltimore County are displayed separately from Milwaukee/New York, with only descriptive statistics presented for Baltimore County. (Appendix D contains performance characteristics for each site and all sites combined for each protocol.)

5.2.2 Findings based on Dwellings from Milwaukee/New York (pre-1950 dwellings)

5.2.2.1 General Findings

One hundred thirty-two dwellings (63 from Milwaukee and 69 New York City) met the eligibility criteria for inclusion in the analysis of pre-1950 dwellings. The geometric mean blood lead level for the EBL and non-EBL children in these dwellings was 15 and 4 µg/dL, respectively (Table 5.3). The geometric mean of the household average dust lead loading was 10 $\mu g/ft^2$ on floors (including carpets) and 100 $\mu g/ft^2$ on window sills. The geometric mean soil lead concentration was 320 ppm at play areas and 1,520 ppm at perimeters. For all types of hazards, the "average" levels of lead fell below the most recent standards[4] set by HUD and EPA.

The occurrence of lead-based paint hazards by each testing component is summarized in Table 5.3. Under the lead-based paint hazards definitions published by EPA in its final Section 403 rules,

- 11% of dwellings failed based on a mean floor dust lead loading ≥ 40 $\mu g/ft^2$
- 20% of dwellings failed based on a mean window sill dust lead loading ≥ 250 $\mu g/ft^2$,
- 10% of dwellings failed based on a play area soil lead sample ≥ 400 ppm,
- 34% of dwellings failed based on a perimeter soil lead sample ≥ 1,200 ppm,
- 99% of dwellings failed based on at least one testing combination with non-intact, lead-based paint (≥ 1 mg/cm^2)

These results are based on household average dust lead loadings as called for in the regulations. Using household maximums for dust (the HUD 1012/1013 standards), 27 percent of dwellings would have failed based on floor dust and 21 percent of dwellings would have failed based on window sill dust lead. These results can be compared to the incidence of children with elevated blood lead levels (50%).

5.2.2.2 Risk Assessment Performance Characteristics

The performance characteristics of the risk assessment protocols as defined in Section 5.1.2.1 are presented in Table 5.4. The risk assessments all performed in a similar manner. Between 77 and 83 percent of the dwellings failed the risk assessments while 50% of the dwellings housed children with elevated blood lead levels. Unfortunately, none of the risk assessment protocols were associated with blood lead status. The lowest p-value among the four protocols tested was 0.822, well higher than the 0.05 level that would demonstrate predictive power. None of the risk assessment protocols were successful in discriminating between the study units that housed lead-poisoned children and those that did not.

---

[4] The EPA 403 regulations define soil hazards at two locations: play area and rest-of-yard (perimeter plus non-play area sample). Because is not clear from EPA's interpretive guidance whether the curb sample collected in this study would qualify as a valid non-play area sample, the perimeter sample and a "Yard" (perimeter and curb) are both presented on Table 5.3. As the greater value, the perimeter result is compared to the EPA standard.

There was little difference between the performance characteristics of the four protocols. The sensitivity of tests averaged 81 percent, while the specificity averaged 19 percent. These protocols have been much more successful in identifying homes with lead-poisoned children than in identifying "low-risk" homes. The positive and negative predictive values fluctuate around 50 percent. All of the tests fell well below the EPA's target negative predictive value of 95 percent discussed in EPA's proposed Section 403 rules.

Table 5.3: Geometric Means, 95% Confidence Intervals and Failure of Environmental Assessments in Milwaukee and New York Combined (N=133)

Component	Geometric Mean (95% CI)	Standard	%Dwellings where maximum fails	%Dwellings where average fails
Floors	10 (1-120) µg/ft^2	100 µg/ft^2	11	5
		40 µg/ft^2	27	11
Sills	100 (3-2,929) µg/ft^2	500 µg/ft^2	15	12
		250 µg/ft^2	21	20
Troughs	1,880 (12-296,825) µg/ft^2	800 µg/ft^2	43	41
Play Area Soil	320 (46-2,239) ppm	400 PPM	10	
Yard[1] Soil	974 (130-7,303) ppm	2000 PPM	16	
		1200 PPM	27	
Perimeter Soil	1,520 (176-13,118) ppm	2000 PPM	24	
		1200 PPM	34	
Poor LBP[2]		1 mg/cm^2	74	
Fair or Poor LBP[2]		1 mg/cm^2	99	
Blood Lead	EBL 15 (7,31) Not EBL 4 (2,11) µg/dL			

[1] Average of perimeter and curbside soil samples.

[2] Interior of unit or exterior of building (including common exterior)

Note: If soil is completely covered, it does not fail.

Note: soil average includes NON completely covered soil.

Table 5.4: Performance Characteristics for Environmental Testing Protocols Studied[1]

Assessment	Number of Units	Percent Failure	Test of Independence P-Value	Performance Characteristic (95% CI)			
				Sensitivity	Specificity	Positive Predictive Value	Negative Predictive value
Risk Assessment							
HUD Guidelines (Average dust)	132	82	0.822	81 (69, 89)	17 (9, 28)	50 (40, 60)	46 (26, 67)
HUD Guidelines (Maximum dust)	132	83	1.0	82 (71, 90)	17 (9, 28)	50 (41, 60)	48 (27, 69)
HUD Interim (Maximum dust)	132	80	1.0	79 (67, 88)	20 (11, 32)	50 (41, 60)	48 (29, 68)
EPA Current (Average dust)	132	80	1.0	81 (69, 89)	20 (11, 32)	51 (41, 61)	50 (30, 70)
Lead Hazard Screen							
HUD Original	53	74	1.0	73 (50,89)	26 (12,45)	41 (26,58)	57 (29,82)
EPA Current	53	55	0.588	50 (28,72)	42 (25,61)	38 (21,58)	54 (33,74)
HUD Current	53	68	0.134	55 (32,76)	23 (10,41)	33 (19,51)	41 (18,67)

[1] See Appendix Tables D1 and D2 for more extensive details.

It should be noted that the positive and negative predictive values will change based the percentage of children with an elevated blood lead level in a population. Because the general population of children in this country has a much lower prevalence rate of elevated blood lead levels than 50 percent, the positive predictive value would decrease in the general population while the negative predictive value would increase. Sensitivity and specificity are less likely to change based on the population, assuming that the study population is reasonably representative of pre-1950 homes.

Although none of the risk assessment protocols demonstrated a relationship between the result of the diagnostic test and the blood lead outcomes, further investigation later in this report will provide information about potential protocols that would have been predictive in this study's pre-1950, urban population. Later in this section, the impact of using a more prescriptive set of dust sampling locations are described.

5.2.2.2.1 Impact of the Revised Definition of Deteriorated Lead-Based Paint

HUD's 1012/1013 rules and EPA's 403 rules changed the definition of deteriorated leaded paint from lead-based paint in "poor" condition (see Section 5.1.3.2) to non-intact lead-based paint. When the new definition was used in the analysis for **EPA Current** risk assessment protocols, the performance characteristics changed substantially. The number of dwellings failing the test increased from 80 to 99 percent. The sensitivity rose from 81 to 100 percent while the specificity dropped from 20 to 2 percent. The positive predictive value remained the same (51%) while the negative predictive power increased from 50 to 100 percent. However, the test of independence continued to find no significant relationship between the environmental assessment and the presence or absence of a lead-poisoned child. With a failure rate of 99 percent, it appears self-evident that an owner of a pre-1950 property would do just as well to assume the dwelling is high-risk rather than pay the cost of a risk assessment, if the owner hoped to predict risk. (A property owner might still hire a risk assessor to develop a lead hazard control plan.)

5.2.2.2.2 Impact of Prescribing the Dust Sampling Locations

The analytical plan for this study was designed to maintain as many of the rules of the original risk assessment protocols as feasible. A central theme in both the *Guidelines* and in the final Section 402/404 rules from EPA is that the risk assessor is expected to use much discretion when selecting dust sampling locations. While the risk assessor is expected to sample locations where children under the age of six are likely to be exposed, the number and location (either within a dwelling or within a room) of the samples has never been prescribed. Since this was a central premise for all of the protocols, it was maintained when examining performance in this study.

Given the poor predictive value of the protocols in this study, an obvious question is whether the problem lies with the protocols or the risk assessor's professional judgment. To test this question, the performance of the risk assessment protocols when the risk assessor chose the locations was compared with the performance when researchers selected alternative locations.

The risk assessors selected for this study, who were all certified and had demonstrated some experience in the field, stated that they would have selected the sampling locations shown in Tables 5.5 through 5.7, based on their professional judgment. Of these locations, risk assessors preferred to sample floors rather than windows, uncarpeted floors rather than carpeted floors, and bedrooms, living rooms, and kitchens in that order. On floors, risk assessors preferred central floor samples. On windows, risk assessors were fairly evenly split between window sills and window troughs. In 23 percent of the dwellings, the risk assessor could not or chose not to sample any window sampling locations. The sample location decisions that the risk assessors made generally seem to conform to guidance in the *Guidelines*.

Table 5.5: Surface Types Chosen by Risk Assessors at Floor, Window Sill, and Window Trough Locations

Total Dust Samples Chosen in Dwelling	Number of Dwellings	Average Number of Dust Samples For Each Surface Type		
		Floor (Uncarpeted/Carpeted)	Window Sill	Window Trough
6	31	4.2 (2.2/2.0)	0.7	1.1
7	21	4.8 (3.0/1.8)	0.6	1.7
8	81	5.9 (4.4/1.5)	1.1	1.1
7.4 (Average)	133	5.3 (3.6/1.7)	0.9	1.2

Table 5.6: Room Locations Chosen by Risk Assessors

Room Location	Number of Dust Samples Chosen for Each Surface Type		
	Floors	Window Sills	Window Troughs
Unit Entryway	55	-	-
Bedroom (1)	*222*	*43*	*56*
Bedroom (2)	*26*	*6*	*13*
Any Bedroom	248	49	69
Living Room	215	39	44
Kitchen	133	25	29
Bathroom	31	4	3
Other	24	4	9
Total	**706**	**121**	**154**

Note: Some floor totals are greater than 133 because risk assessors may have identified more than one floor sampling location in a room.

Bedroom (1) was the enrolled child's bedroom. Bedroom (2) was the bedroom of another child.

Table 5.7: Floor Locations Chosen by Risk Assessors

Floor Location	Frequency
Central Floor	296
Entry Floor	166
Window Floor	146
Perimeter Floor	98

The study attempted to determine whether the risk assessors' decisions regarding which sampling locations might have adversely affected performance of the risk assessment protocols. Instead of using the 6-8 dust sampling locations selected by each risk assessor, a standard set of dust sampling locations was analyzed. The sampling locations selected for the statistical analyses in this section were solely based on the judgment of the researchers.

The standard set of locations selected were central floor samples collected from four locations: entryway, living room, the index child's bedroom (bedroom 1), and kitchen. Both uncarpeted floors and carpeted floors were acceptable. Window sill samples from two locations (living room and kitchen) and window trough samples from two locations (living room and bedroom) were also chosen. If the risk assessor did not collect the sample or there was no standard for the sample type, no alternative sampling site was selected.

For this analysis, only the performance characteristics for the **HUD Guidelines-Maximum, HUD Guidelines -Average** and **HUD Interim** were examined. Analytical results suggest that implementing standard sampling locations would have little impact (Table 5.8). The tests of independence continued to find that the risk assessments have no significant relationship to the blood lead outcomes. The use of a standard set of sampling locations (instead of the locations selected by the risk assessor) resulted in slightly higher failure rates (82% to 84%, on average) and sensitivities (81% to 83%), and slightly lower specificities (18% to 15%). Since there was no change in the predictive power and the performance characteristics are similar, the results support allowing risk assessors to use their professional judgment, at least under the current risk assessment protocols.

Table 5.8: Comparison of Performance Characteristics for Risk Assessment Protocols (Dust Sampling Locations Selected by Risk Assessor v. Locations Selected by Study)[1]

Assessment	Selector of Dust Sample Location	Percent Failure	Test of Independence P-Value	Performance Characteristics (95% CI)			
				Sensitivity	Specificity	Positive Predictive Value	Negative Predictive value
Risk Assessment							
HUD Guidelines (Average dust)	Risk Assessor	82	0.822	81 (69,89)	17 (9,28)	50 (40,60)	46 (26,67)
	Study	84	0.635	82 (71,90)	14 (6,24)	49 (40,59)	43 (22,66)
HUD Guidelines (Maximum dust)	Risk Assessor	83	1.000	82 (71,90)	17 (9,28)	50 (41,60)	48 (27,69)
	Study	85	0.809	84 (73,92)	14 (6,24)	50 (40,59)	45 (23,68)
HUD Interim (Maximum dust)	Risk Assessor	67	1.000	79 (67,88)	20 (11,32)	50 (41,60)	48 (29,68)
	Study	83	1.000	82 (71,90)	17 (9,28)	50 (41,60)	48 (27,69)

[1]See Appendix Tables D1 and D3 for more extensive details.

5.2.2.3 Lead Hazard Screen Performance Characteristics

The performance characteristics of the three lead hazard screen protocols as defined in Section 5.1.2.2 are presented in Table 5.4. The performance of the three lead hazard screen protocols varied more from protocol to protocol than the performance of the risk assessment protocols. Seventy-four percent of the dwellings assessed using the **HUD Original Screen** protocols failed while only 55 percent of the dwellings assessed using the **EPA Current Screen** protocols failed. The **HUD Original Screen** had a higher sensitivity (73% v. 50%), but a lower specificity (26% v. 42%) than the **EPA Current Screen.** The difference solely reflected the change from trough sampling to sill sampling. By all measures, the **HUD Current Screen** displayed poorer performance characteristics by all measures than the HUD Original Screen. The failure rates of the dwellings undergoing the assessments (between 55 and 74%) compare to 42 percent of the children in those dwellings having elevated blood lead levels.

As with the risk assessment protocols, the test of independence found no significant relationship between the lead hazard screen protocols and the blood lead outcomes (EBL/no EBL). However, this finding is of less concern for the screen than for the risk assessment protocols, because the principal goal of the screen is to identify only "low-risk" dwellings. A screen protocol with a high sensitivity and high negative predictive value is preferable in order to reduce the chance of "high-risk" homes slipping through the screen. At the same time, however, a screen protocol that results in a high percentage of dwellings failing the screen when compared to the prevalence of lead poisoning in the target population is unlikely to pass a cost-benefit analysis.

Given these goals, the current screen protocols did not perform well in this pre-1950, urban study population. While the definition of "high" sensitivity and "high" negative predictive value is open to discussion, if a reasonable criterion was at least 80%, then none of the lead hazard screen protocols achieved that standard. The protocol coming closest, the **HUD Original Screen**, had a sensitivity of 73% and a negative predictive value of 57%, but also had the highest rate of failure (74%).

The performance characteristics for the lead hazard screen protocols when applied to pre-1950 housing are supportive of both HUD and EPA's guidance to avoid using screens in pre-1960 properties. Although all of the dwellings investigated were in "good" condition, the prevalence of lead-poisoned children was high (42%) and the failure rates were all at least 55 percent. For the majority of property owners in this population of housing, there would be additional costs and no benefit to conducting the screen.

5.2.3 Findings Based on Dwelling from Baltimore County (post-1950 dwellings)

5.2.3.1 General Findings

Ninety-seven dwellings from Baltimore County met the eligibility criteria for inclusion in the analysis of post-1950 dwellings. However, as discussed in section 5.2.1, the analyses in this section of the report are limited to descriptive statistics. With a prevalence of lead-poisoned children of only one percent in this study population, comparative statistics between homes with and without children with elevated blood leads are not appropriate.

The geometric mean blood lead level for the enrolled children in these dwellings was 3 µg/dL (Table 5.9). The geometric mean of the household average dust lead loading was 3 $\mu g/ft^2$ on floors (including carpets) and 27 $\mu g/ft^2$ on window sills. The geometric mean soil lead concentration was 45 ppm at play areas and 87 ppm at perimeters. In all cases, the "average" levels of lead fell below the most recent standards set by HUD and/or EPA and were well below the pre-1950 levels described in Section 5.2.2.1.

The occurrence of lead-based paint hazards by component is summarized in Table 5.8. Using the lead-based paint hazard definitions published by EPA in its final Section 403 rules,

- 0% of dwellings failed based on a mean floor dust lead loading ≥ 40 $\mu g/ft^2$
- 3% of dwellings failed based on a mean window sill dust lead loading ≥ 250 $\mu g/ft^2$,
- 1% of dwellings failed based on a play area soil lead sample ≥ 400 ppm,
- 1% of dwellings failed based on a perimeter soil lead sample ≥ 1,200 ppm,
- 81% of dwellings failed based on at least one testing combination with non-intact, lead-based paint (≥ 1 mg/cm^2)

These results are based on household average dust lead loadings as called for in the regulations. Using household maximums for dust, no floor dust results would have failed and 10 percent of dwellings would have failed based on window sill dust lead.

Individually, the percentage of failures corresponds fairly well with the prevalence of lead-poisoning, with the exception of maximum window sill dust and non-intact deteriorated lead-based paint. Had the previous standard for deterioration ("poor" condition) been considered, the percentage of dwellings failing the paint assessment would have still been relatively high (47 percent).

Table 5.9: Geometric Means, 95% Confidence Intervals and Failure of Environmental Assessments in Baltimore County (N=97)

Component	Geometric Mean (95% CI)	Standard	%Dwellings where maximum fails	%Dwellings where average fails
Floors	3 (1-7) µg/ft^2	100 µg/ft^2	0	0
		40 µg/ft^2	0	0
Sills	27 (3-270) µg/ft^2	500 µg/ft^2	3	1
		250 µg/ft^2	10	3
Troughs	339 (11-10,487) µg/ft^2	800 µg/ft^2	31	28
Play Area Soil	45 (11-191) PPM	400 PPM	1	
Yard[1] Soil	88 (26-289) PPM	2000 PPM	0	
		1200 PPM	0	
Perimeter Soil	87 (18-426) PPM	2000 PPM	1	
		1200 PPM	1	
Poor LBP[2]		1 mg/cm^2	47	
Fair or Poor LBP[2]		1 mg/cm^2	81	
Blood Lead	3 (1-6) µg/dL			

[1] Average of perimeter and curbside soil samples.
[2] Interior of unit or exterior of building (including common exterior)
Note: If soil is completely covered, it does not fail. The soil average includes only NON-completely covered soil.

5.2.3.2 Risk Assessment Findings

As in the pre-1950 properties, the risk assessment protocols all performed in a similar manner (Table 5.10). Between 48 and 57 percent of the dwellings failed the risk assessments (when using a paint condition of poor as the paint assessment standard). There was not a significant difference in the geometric mean blood lead in homes that passed and failed the risk assessment. The paint assessment was by far the main reason for dwellings to fail the risk assessment protocols in Baltimore County. Under the **EPA Current** protocols using no *de minimus* level for paint deterioration, 81 percent of the dwellings in Baltimore County would have failed the risk assessment.

Although all of the risk assessment protocols identified approximately half of the houses as being "high-risk", none of the protocols identified the home of the one EBL child as being "high-risk." While this study did not intend to identify the exposure sources for each poisoned child in the study, and has not attempted to identify the sources for the child in Baltimore County, this child may be illustrative of the fact that not all poisoned children receive significant exposure from the dwelling. Children can come in contact with lead at other residences and day-care facilities and from non-housing sources such as pottery and traditional medicines. No matter how successful a risk assessment protocol is at identifying dwelling unit sources of exposure, no protocol should be expected to be perfect.

Table 5.10: Percent Failure Environmental Testing Protocols for Baltimore Post-1950 Dwellings (N=97) [1]

Protocol	Percent Failure
HUD Guidelines (Average dust)	56
HUD Guidelines (Max. dust)	57
HUD Interim (Maximum dust)	53
EPA Current (Average dust)	48

[1] See Appendix Table D1 for more extensive details.

5.2.3.3 Lead Hazard Screen Findings

Fifty-eight dwelling units in Baltimore County met the eligibility criteria to be part of the analysis of lead hazard screens. Even with the tighter dust standards in place, none of the three screen protocols were able to identify the dwelling that housed the lead-poisoned child. Failure rates were 50 percent, 36 percent, and 31 percent for the **HUD Original Screen**, **HUD Current Screen** and **EPA Current Screen**, respectively.

Arguably, the protocols were not applied to the appropriate subset of housing since 92 percent of the Baltimore County properties in the study were built prior to 1960. However, given the low incidence of children with elevated blood lead levels and the low levels of lead-in-dust and lead-in-soil found at these properties, it would be hard to identify a better and more appropriate set of properties that should be eligible for a screen. See Appendix Table D2 for more extensive details.

6.0 EXPLORATION OF INDIVIDUAL FACTORS INFLUENCING RISK ASSESSMENT PERFORMANCE (OBJECTIVES #3-7)

When developing the statistical analysis plan for this study, the researchers recognized that a study designed only to answer the question of whether the current risk assessment protocols are valid would likely create more questions than it answered. If the current protocols were demonstrated to be valid, then questions would remain about whether the scope of the protocols could be reduced (i.e., conducted at less cost) and remain valid. If the current protocols were not demonstrated to be valid, then it would be important to explore why and demonstrate that a valid protocol is possible. Therefore, this study was designed to look at the individual components of an environmental assessment and then use the knowledge gained from that exploration to identify better options (Section 7).

In this section, the influence of dust and paint assessments on overall risk assessment performance is considered, as is the influence of friction/impact surfaces on dust lead loadings. The assessment of dust lead is examined by surface type (uncarpeted floor, carpet, window sill, and window trough) and by sampling location. The influence of the variability of dust lead loadings is also considered. While there are many other components of the environmental assessment that could be considered (e.g., the influence of soil lead levels and soil cover), the five factors chosen are considered most critical to understanding how to improve the current protocols.

6.1 Optimum Standards for the Maximum and Average of Dust Wipe Samples

6.1.1 Purpose

A number of previous studies have documented a significant relationship between dust lead and blood lead [10,11]. This study was not designed to replicate this earlier research, but instead considered questions that have been raised as the research has been incorporated into the regulatory process. To better refine the risk assessment protocols, answers are needed to questions such as:

- Is the average or the household maximum the optimal measure of household lead dust when predicting high- and low-risk dwellings?
- If both measures are reliable, should the standard for the average differ from that of the maximum?
- Is a composite sample truly equivalent to the average of individual samples, and can the mean standard and composite standard be the same?

Answers to these questions will help determine whether the average dust lead level or the maximum dust lead level optimizes the predictive power of a risk assessment. They will also help determine whether separate dust lead standards are needed for composite samples versus single surface samples.

Of course, underlying these applied research questions is the fundamental question of how well household dust lead relates to blood lead outcomes. As discussed in Section 5, current risk assessment protocols are not significant predictors of “high-risk” dwellings.

This objective will explore whether dust lead loadings by themselves are significant predictors of blood lead outcomes and whether the predictive power of the current protocols can be improved.

6.1.2 Objectives 3a&b: Describe the performance characteristics for floors, window sills and troughs for the average and the maximum of single-surface samples.

6.1.2.1 Methodology

Using the performance characteristics analyses described in Section 5.1.1, the utility of the dust lead sampling portion of the risk assessment protocols were examined. The performance characteristics of the dust lead were considered in three separate stages:

1. The performance characteristics of dust lead were assessed by individual sample type (floors (all, uncarpeted only, and carpeted only), window sills and window troughs) using the 1995 EPA/HUD interim risk assessment standards (100, 500, 800), 1999 HUD and 2001 EPA final risk assessment standards (40, 250) and the final HUD Screen standards (25, 125).
2. The performance characteristics of the dust lead protocols of the three current risk assessment protocols and **HUD Current Screen** protocols (defined in section 5.1.2.1) as well as the **HUD Interim** protocols were assessed using the household maximum and average dust lead loadings. For each protocol, the dust lead standard dictated by that protocol was used. All three floor surface types were considered (all floors, uncarpeted, and carpeted), but only the findings for all floor samples are presented below. The other findings are presented in Appendix D.
3. Any dust lead protocol that would be an improvement over the current protocols was identified. Better protocols were defined as protocols that were at least marginally significant predictors of blood lead outcomes ($p<0.1$).

The first two stages were intended to be informative of the performance of the current protocols yet be comparable to the protocols examined in the final stage. Therefore, a decision was made to use the fixed sample locations identified in Section 5.2.2.2.2. Up to four floor samples and up to two window sill and trough samples were selected from the Unit Entry, Living Room, Kitchen and Index Child's Bedroom. On non-Entry floors, the central location was selected.

In Section 5, the influence of site (Baltimore County, Milwaukee, New York City) on an evaluation of the complete risk assessment protocols was discussed. Similar issues arose when examining the dust lead protocols separately from the other components of a risk assessment. When all sites are combined, the dust protocols were often statistically significant predictors of blood lead outcome (EBL/non-EBL). However, the results are dependent on site and when considered separately, the results for each site are often quite different. While the full set of performance analyses are presented in Appendix D, the results presented below are limited to the pre-1950 dwellings in Milwaukee and New York. The results of the post-1950 dwellings (Baltimore County) are not presented, because findings based on a single lead-poisoned child out of 100 dwellings would not be informative. These results are presented in Appendix Tables D8 and D9.

In Section 7, evidence is presented that suggests that the relationship between environmental lead at the enrolled child's home and the child's blood lead status (EBL/Non-EBL) was very different in Milwaukee from New York City. When data from the two sites were combined, lead-based paint risk assessments were at best marginally significantly associated with blood lead levels. However, by separating the two sites, assessments of the environmental lead media at a child's home were strongly related with blood lead status in Milwaukee. For this objective, Milwaukee and New York City results are presented separately.

6.1.2.2 Findings

6.1.2.2.1 Performance Characteristics of Dust Lead for Individual Surface Types

Of the 64 dwellings in Milwaukee that met the basic study eligibility criteria, all dwellings had at least one sample available on all floors, uncarpeted floors, window sills and window troughs. Fifty-four (54) dwellings had at least one carpeted floor sample. In New York City, all 69 eligible dwellings had at least one sample available on all floors, uncarpeted floors, window sills and window troughs. Thirty-two (32) dwellings in New York had at least one carpeted floor sample. An assessment of the performance characteristics of these surface types was conducted using the 1995 EPA/HUD interim dust hazard standards for risk assessments, the 1999 HUD/2001 EPA final standards for risk assessments and the final HUD Screen standards (Tables 6.1.1a-b and 6.1.2a-b).

The analysis results identify four surface/standard combinations that were at least marginally statistically significant predictors of blood lead outcomes (EBL/no EBL) in this pre-1950 study population in Milwaukee using the current or interim standards:

All floors (max)	25 µg/ft^2	p=0.01
All floors (mean)	25 µg/ft^2	p=0.03
Carpeted floors (max)	25 µg/ft^2	p=0.04
All floors (max)	40 µg/ft^2	p=0.07

None of the current or interim standards were significant predictors of blood lead outcomes in New York City.

A diagnostic test in pre-1950 Milwaukee dwellings using a standard of 25 µg/ft^2 on all floors resulted in a fairly good balance of performance characteristics when the maximum floor dust lead loading was used (sensitivity=71%, specificity=61%). The results indicate that floor dust lead, when measured alone using the lead hazard screen standard, did a fairly good job of identifying both "low- risk" and "high risk" dwellings.

Table 6.1.1a: Performance Characteristics for Individual Dust Lead Sample Types Using the Maximum Dwelling Unit Dust Lead Loading (Milwaukee) [1]

Assessment		Number of Units	Percent Failure	Test of Independence P-Value	Performance Characteristic (95% CI)			
					Sensitivity	Specificity	Positive Predictive Value	Negative Predictive value
Sample Type/Standard (μg/ft2)								
All Floors	100	64	14	1.000	13 (4,30)	85 (68,95)	44 (14,79)	51 (37,65)
	40		39	0.072*	52 (33,70)	73 (54,87)	64 (43,82)	62 (45,77)
	25		55	0.014**	71 (52,86)	61 (42,77)	63 (45,79)	69 (49,85)
Uncarpeted Floors	100	64	8	0.667	10 (2,26)	94 (80,99)	60 (15,95)	53 (39,66)
	40		27	0.159	35 (19,55)	82 (65,93)	65 (38,86)	57 (42,72)
	25		39	0.200	48 (30,67)	70 (51,84)	60 (39,79)	59 (42,74)
Carpeted Floors	100	54	11	0.675	8 (1, 26)	86 (68,96)	33 (4,78)	52 (37,67)
	40		22	0.188	32 (15,54)	86 68,96)	67 (35,90)	60 (43,74)
	25		26	0.035**	40 (21,61)	86 (68,96)	71 (42,92)	63 (46,77)
Window Sills	500	64	47	0.465	42 (25,61)	48 (31,66)	43 (25,63)	47 (30,65)
	250		59	0.803	61 (42,78)	42 (25,61)	50 (33,67)	54 (33,73)
	125		77	0.242	84 (66,95)	30 (16,49)	53 (38,67)	67 (38,88)
Window Troughs	800	64	81	1.000	81 (63,93)	18 (7, 35)	48 (34,62)	50 (21,79)
	400		86	0.729	84 (66,95)	12 (3,28)	47 (34,61)	44 (14,79)

*=$p<0.10$, **=$p<0.05$

[1] See Appendix Table D8 for more extensive details.

Table 6.1.1b: Performance Characteristics for Individual Dust Lead Sample Types Using the Maximum Dwelling Unit Dust Lead Loading (New York City) [1]

Assessment		**Number of Units**	**Percent Failure**	**Test of Independence P-Value**	**Performance Characteristic (95% CI)**			
					Sensitivity	**Specificity**	**Positive Predictive Value**	**Negative Predictive value**
Sample Type/Standard (µg/ft2)								
All Floors	100	69	1	1.000	3 (0,15)	100 (89,100)	100 (3,100)	49 (36,61)
	40		6	0.615	8 (2,22)	97 (84,100)	75 (19,99)	49 (37,62)
	25		9	0.675	11 (3,26)	94 (80,99)	67 (22,96)	49 (36,62)
Uncarpeted Floors	100	69	1	1.000	3 (0,15)	100 (89,100)	100 (3,100)	49 (36,61)
	40		4	1.000	6 (1,19)	97 (84,100)	67 (9,99)	48 (36,61)
	25		7	1.000	8 (2,22)	94 (80,99)	60 (15,95)	48 (36,61)
Carpeted Floors	100	32	0	-	-	-	-	-
	40		3	1.000	6 (0,29)	100 (78,100)	100 (3,100)	48 (30,67)
	25		3	1.000	6 (0,29)	100 (78,100)	100 (3,100)	48 (30,67)
Window Sills	500	69	4	0.603	3 (0,15)	94 (80,99)	33 (1,91)	47 (35,60)
	250		16	0.330	11 (3,26)	79 (61,91)	36 (11,69)	45 (32,58)
	125		19	0.761	17 (6,33)	79 (61,91)	46 (19,75)	46 (33,60)
Window Troughs	800	69	22	0.384	17 (6,33)	73 (54,87)	40 (29,68)	44 (31,59)
	400		39	0.629	36 (21,54)	58 (39,75)	48 (29,68)	45 (30,61)

*=p<0.10, **=p<0.05

[1] See Appendix Table D8 for more extensive details.

Table 6.1.2a: Performance Characteristics for Individual Dust Lead Sample Types Using the Arithmetic Mean Dwelling Unit Dust Lead Loading (Milwaukee) [1]

Assessment		**Number of Units**	**Percent Failure**	**Test of Independence P-Value**	**Performance Characteristic (95% CI)**			
					Sensitivity	**Specificity**	**Positive Predictive Value**	**Negative Predictive value**
Sample Type/Standard (µg/ft2)								
All Floors	100	64	5	0.607	6 (1,21)	97 (84,100)	67 (9,99)	52 (39,65)
	40		16	0.178	23 (10,41)	91 (76,98)	70 (35,93)	56 (41,69)
	25		31	0.030**	45 (27,64)	82 (65,93)	70 (46,88)	61 (45,76)
Uncarpeted Floors	100	64	5	0.607	6 (1,21)	97 (84,100)	67 (9,99)	52 (39,65)
	40		19	0.208	26 (12,45)	88 (72,97)	67 (35,90)	56 (41,70)
	25		30	0.173	39 (22,58)	79 (61,91)	63 (38,84)	58 (42,72)
Carpeted Floors	100	54	4	0.493	None	93 (77,99)	None	52 (38,66)
	40		17	0.718	20 (7,41)	86 68,96)	56 (21,86)	56 (40,70)
	25		22	0.188	32 (15,54)	86 (68,96)	67 (35,90)	60 (43,74)
Window Sills	500	64	31	0.791	29 (14,48)	67 (48,82)	45 (23,68)	50 (31,65)
	250		50	1.000	48 (30,67)	48 (31,66)	47 (29,65)	50 (32,68)
	125		72	0.784	74 (55,88)	30 (16,49)	50 (35,65)	56 (31,78)
Window Troughs	800	64	80	1.000	81 (63,93)	21 (9,39)	49 (35,63)	54 (25,81)
	400		84	1.000	84 (66,95)	15 (5,32)	48 (34,62)	50 (19,81)

*=p<0.10, **=p<0.05

[1] See Appendix Table D9 for more extensive details.

Table 6.1.2b: Performance Characteristics for Individual Dust Lead Sample Types Using the Arithmetic Mean Dwelling Unit Dust Lead Loading (New York City) [1]

Assessment		Number of Units	Percent Failure	Test of Independence P-Value	Performance Characteristic (95% CI)			
					Sensitivity	Specificity	Positive Predictive Value	Negative Predictive value
Sample Type/Standard (μg/ft2)								
All Floors	100	69	0	-	-	-	-	-
	40		3	0.494	6 (1,19)	100 (89,100)	100 (16,100)	49 (37,62)
	25		3	0.494	6 (1,19)	100 (89,100)	100 (16,100)	49 (37,62)
Uncarpeted Floors	100	69	0	-	-	-	-	-
	40		3	0.494	6 (1,19)	100 (89,100)	100 (16,100)	49 (37,62)
	25		3	0.494	6 (1,19)	100 (89,100)	100 (16,100)	49 (37,62)
Carpeted Floors	100	32	0	-	-	-	-	-
	40		0	-	-	-	-	-
	25		0	-	-	-	-	-
Window Sills	500	69	4	0.603	3 (0,15)	94 (80,99)	33 (1,91)	47 (35,60)
	250		13	0.728	11 (3,26)	85 (68,95)	44 (14,79)	47 (34,60)
	125		17	0.531	14 (5,29)	79 (61,91)	42 (15,72)	46 (32,59)
Window Troughs	800	69	17	0.531	14 (5,29)	79 (45,90)	42 (15,72)	46 (32,59)
	400		33	0.621	31 (16,48)	64 (45,80)	48 (27,69)	46 (32,59)

*=p<0.10, **=p<0.05

[1] See Appendix Table D9 for more extensive details.

6.1.2.2.2 Performance Characteristics of Dust Lead Protocols for Sample Types Considered Simultaneously

For the six dust sampling protocols examined, one assessment was significantly related to the blood lead outcomes in the pre-1950 dwellings of Milwaukee: the current HUD Screen standards using the maximum dust lead level in the home to identify hazards (Tables 6.1.3a and 6.1.3b). In Milwaukee, this protocol had a sensitivity of 94 percent and a specificity of 27 percent. No dust sampling protocol was predictive of blood lead outcomes in New York City. In New York, the current HUD Screen standard had a sensitivity and specificity of 25 percent and 76 percent, respectively, but had no statistical association (p=1.0).

6.1.2.2.3 Examination of an Improved Dust Sampling Plan

The first two stages of this analysis identified five protocols for dust collection in Milwaukee that were significantly related to the presence or absence of a child with an elevated blood lead level. These findings offer support to the hypothesis that residential dust lead loadings were a lead exposure source for children in Milwaukee. However, the lack of predictive power for many of the current protocols raised questions about whether alternative standards, especially for window sills and troughs, might improve the predictive power of lead risk assessments.

To establish alternative dust sampling combinations, an *a priori* list of alternative dust lead standards was developed for each sample type. The list included nine potential dust standards for floor dust lead (5, 10, 15, 20, 25, 30, 35, 40 $\mu g/ft^2$ and none); ten potential dust standards for window sill dust lead (50, 100, 150, 200, 250, 300, 350, 450, 500 $\mu g/ft^2$ and none); and five potential dust standards for window trough dust lead (400, 800, 5,000, 10,000 $\mu g/ft^2$ and none). Every permutation of sample type and standard was then tested for a possible relationship with the blood lead outcome for a dwelling. Using the fixed sampling scheme described in Section 5.2.2.2.2, samples from the unit entry, living room, child's bedroom and kitchen were analyzed. Floor samples were analyzed using the central floor sample. A separate set of analyses were conducted excluding the dwelling unit entry floor sample.

Because the current regulations call for the household mean dust lead levels to be used for risk assessments, a decision was made to limit these analyses to mean values for each sample type (floors, window sills and window troughs). It is recognized that in the two earlier stages, more protocols that used the maximum household dust lead value were predictive of blood lead status than protocols that used the household mean dust lead loading. However, results presented in Section 6.1.3 suggest that this may be due to the current standards being higher than what would be optimal when using the household mean.

Table 6.1.3a: Performance Characteristics for Dust Lead Assessment Alternatives Using All Floor Surface Types (Carpeted and Uncarpeted) (Milwaukee) [1]

Assessment	Number of Units	Percent Failure	Test of Independence P-Value	Performance Characteristic (95% CI)			
				Sensitivity	Specificity	Positive Predictive Value	Negative Predictive value
Dust Protocol/Standards (μg/ft2)							
HUD Guidelines-Max (f-100, s-500, t-800)	64	88	0.468	84 (66,95)	9 (2,24)	46 (33,60)	38 (9,76)
HUD Guidelines-Avg. (f-100, s-500, t-800)	64	83	1.000	84 (66,95)	18 (7,35)	49 (35,63)	55 (23,83)
HUD Interim-Max (f-40, s-250)	64	72	0.169	81 (63,93)	36 (20,55)	54 (39,69)	67 (41,87)
HUD Interim-Avg. (f-40, s-250)	64	52	1.000	52 (33,70)	48 (31,66)	48 (31,66)	52 (33,70)
HUD Current Screen-Max (f-25, s-125)	64	83	0.045**	94 (79,99)	27 (13,46)	55 (40,68)	82 (48,98)
HUD Current Screen-Avg. (f-25, s-125)	64	81	1.000	90 (74,98)	27 (13,46)	54 (39,68)	75 (43,95)

[1] See Appendix Tables D10 and D11 for more extensive details.

Table 6.1.3b: Performance Characteristics for Dust Lead Assessment Alternatives Using All Floor Surface Types (Carpeted and Uncarpeted) (New York City)[1]

Assessment	Number of Units	Percent Failure	Test of Independence P-Value	Performance Characteristic (95% CI)			
				Sensitivity	Specificity	Positive Predictive Value	Negative Predictive value
Dust Protocol/Standards (μg/ft2)							
HUD Guidelines-Max (f-100, s-500, t-800)	69	28	0.419	22 (10,39)	67 (48,82)	42 (20,67)	44 (30,59)
HUD Guidelines-Avg. (f-100, s-500, t-800)	69	22	0.384	17 (6,33)	73 (54,87)	40 (16,68)	44 (31,59)
HUD Interim-Max (f-40, s-250)	69	19	0.761	17 (6,33)	79 (61,91)	46 (19,75)	46 (33,60)
HUD Interim-Avg. (f-40, s-250)	69	16	1.000	17 (6,33)	85 (68,95)	55 (23,83)	48 (35,62)
HUD Current Screen-Max (f-25, s-125)	69	25	1.000	25 (12,42)	76 (58,89)	53 (28,77)	48 (34,62)
HUD Current Screen-Avg. (f-25, s-125)	69	20	1.000	19 (8,36)	79 (61,91)	50 (23,77)	47 (34,61)

[1] See Appendix Tables D10 and D11 for more extensive details.

The performance characteristics for dust sampling combinations that were at least marginally significant ($p<0.10$) were examined. For presentation in this report, the list of dust sampling combinations was then culled to those combinations using the following procedure:

a. For each level of sensitivity, the sampling combination with the highest specificity was selected. If there was >1 protocol with the same sensitivity and specificity, all were selected.
b. For each level of sensitivity and specificity, the sampling combination with the highest level of association with EBL status (i.e., lowest p-value) was selected. If there was >1 protocol with the same sensitivity, specificity, and p-value, all were selected.
c. Finally if *both* the sensitivity *and* the specificity are lower than those for another sampling combination, then the combination was dropped.
d. The list of selected protocols was ranked by sensitivity.

The nine dust sampling combinations that met these criteria (optimal protocols) are presented on Table 6.1.4.

Table 6.1.4: Nine Optimal Dust Sampling Protocols in Milwaukee Study Population

#	**Surface Type(s)**	**With/Without Entry**	**Standard** (µg/ft^2)
1	Floor	No Entry	5
2	Floor, Sill	Entry	10, 250
3	Floor	Entry	10
4a	Floors	No Entry	10
4b	Floor	Entry	15
5	Floor	Entry	20
6	Floor	Entry	25
7	Floor	No Entry	20
8	Floor	No Entry	25

Eight of the nine optimal dust sampling protocols include only floor samples. These protocols had floor dust lead standards that ranged from 5 to 25 µg/ft^2. When the performance characteristics for these protocols were examined using the Milwaukee data, the protocols for the eight floor-only protocols all had levels of significance below 5 percent, and similar Accuracies[5] (63-69%) (Table 6.1.5a). However, the sensitivity-specificity mix ranged widely from 97-30% to 29-97%.

---

[5] Statistical Accuracy is used to measure the probability of making a correct decision. Accuracy is the probability that a dwelling either fails an environmental assessment given that there is a resident child with an elevated blood lead level or passes an environmental assessment given that a resident child does not have an elevated blood lead concentration.

Table 6.1.5a: Performance Characteristics for Nine Optimal Dust Sampling Protocols (Milwaukee) (n=64) [1]

#	Protocol	Percent Failure	Test of Independence P-Value	Performance Characteristic (95% CI)			
				Sensitivity	Specificity	Positive Predictive Value	Negative Predictive value
Dust Protocol/Standards (μg/ft2)							
1	Floor 5	83	0.006	97 (83,100)	30 (16,49)	57 (42,70)	91 (59,100)
2	Floor (w/Entry), Sill 10, 250	80	0.012	94 (79,99)	33 (18,52)	57 (42,71)	85 (55,98)
3	Floor (w/Entry) 10	70	0.001	90 (74,98)	48 (31,66)	62 (47,76)	84 (60,97)
4a	Floors 10	55	0.014	71 (52,86)	61 (42,77)	63 (45,79)	69 (49,85)
4b	Floor (w/Entry) 15	55	0.014	71 (52,86)	61 (42,77)	63 (45,79)	69 (49,85)
5	Floor (w/Entry) 20	44	0.043	58 (39,75)	70 (51,84)	64 (44,81)	64 (46,79)
6	Floor (w/Entry) 25	31	0.030	45 (27,64)	82 (65,93)	70 (46,88)	61 (45,76)
7	Floor 20	20	0.030	32 (17,51)	91 (76,98)	77 (46,95)	59 (44,72)
8	Floor 25	16	0.005	29 (14,48)	97 (84,100)	90 (55,100)	59 (45,72)

[1] See Appendix Tables D12 and D13 for more extensive details.

Table 6.1.5b: Performance Characteristics for Nine Optimal Dust Sampling Protocols (New York) (n=69) [1]

#	Protocol	Percent Failure	Test of Independence P-Value	Performance Characteristic (95% CI)			
				Sensitivity	Specificity	Positive Predictive Value	Negative Predictive value
Dust Protocol/Standards (µg/ft2)							
1	Floor 5	29	0.797	31 (16,48)	73 (54,87)	55 (32,77)	49 (34,64)
2	Floor (w/Entry), Sill 10, 250	22	0.772	19 (8,36)	76 (58,89)	47 (21,73)	46 (33,60)
3	Floor (w/Entry) 10	12	1.000	11 (3,26)	88 (72,97)	50 (16,84)	48 (35,61)
4a	Floors 10	12	0.712	14 (5,29)	91 (76,98)	63 (24,91)	49 (36,62)
4b	Floor (w/Entry) 15	9	1.000	8 (2,22)	91 (76,98)	50 (12,88)	48 (35,61)
5	Floor (w/Entry) 20	3	0.494	6 (1,19)	100 (89,100)	100 (16,100)	49 (37,62)
6	Floor (w/Entry) 25	3	0.494	6 (1,19)	100 (89,100)	100 (16,100)	49 (37,62)
7	Floor 20	3	0.494	6 (1,19)	100 (89,100)	100 (16,100)	49 (37,62)
8	Floor 25	3	0.494	6 (1,19)	100 (89,100)	100 (16,100)	49 (37,62)

[1] See Appendix Tables D12 and D13 for more extensive details.

When floor-only protocols were compared using the same dust lead standard but with and without Unit Entry samples, protocols with the Entry sample had higher sensitivities, but lower specificities. The results also demonstrated that the performance characteristics are equivalent for a mean floor dust lead loading of 10 µg/ft^2 *without* the Entry included as a mean floor dust lead loading of 15 µg/ft^2 *with* the Entry included.

Two dust sampling protocols from New York City were at least marginally predictive (p=0.087) of blood lead outcomes in that study population. These protocols (floors: 10 or 15 µg/ft^2, window sills: 100 µg/ft^2, and window troughs: 400 µg/ft^2) each had a sensitivity of 78 percent and a specificity of 6 percent. However, with both the positive and negative predictive values below 50 percent (47% and 20%, respectively), the results are not indicative of a very useful assessment tool. Furthermore, when these protocols were applied to Milwaukee data, the sampling plans were not associated with blood lead outcomes (p=0.61) and 94 percent of the dwellings would have failed, further suggesting that these standards would not be universally optimal.

For comparison, the performance characteristics for dwellings in New York City are presented for the optimal dust sampling protocols from Milwaukee (Table 6.1.5b). None of the optimal protocols for Milwaukee were significantly associated with blood lead outcomes in New York City. However, if as hypothesized in Section 7, children in New York City were exposed to lead from sources other than the environmental lead in the dwelling unit, these results do not necessarily refute the general power of these protocols as assessment tools. In fact, the high specificities and positive predictive values associated with many of these protocols in New York City suggest that the protocols tend to properly identify homes without children with elevated blood lead levels and homes that should fail the assessment.

6.1.2.3 Discussion of Findings

This study offers evidence that could help identify better dust sampling protocols. First, the findings offer very little evidence to support including either window sill or window trough dust lead samples in a dust lead sampling plan. While previous studies identified a correlation between window dust lead loadings and blood lead levels, no significant relationships between window dust lead and the presence or absence of a child with an elevated blood lead level were identified in either the Milwaukee or New York City study populations.

This is not to suggest that window samples should not be collected at clearance. Other studies have shown that floor dust lead is related to window dust lead. Demonstrating that window dust lead is below clearance levels after treatment could confirm that an indirect source of exposure has been properly addressed in the home.

Second, the findings are supportive of the decision to lower the floor dust lead standards from 100 µg/ft^2. In fact, the results of the analyses suggest that the floor dust lead standards could be further reduced from 40 µg/ft^2. When dust lead samples are considered without the influence of other environmental lead media, the optimal dust lead standards using a household mean were somewhere between 5 and 25 µg/ft^2. (Protocols including other media are examined in Section 7.) While the current floor standard of 40 was marginally significant when the maximum value in the dwelling was used to define a hazard, this standard was not significantly associated with blood lead status when the household mean value was used.

Finally, these results suggest that further consideration should be given to the room locations where dust lead samples are collected. For example, the decision to include or exclude the Unit Entry sample from a household mean affected the protocol's performance characteristics. To optimize predictive power of any given floor dust lead standard, it appears appropriate for regulators to express explicitly whether a sampling plan for a risk assessment should include the Entry. Further discussion about dust sampling locations is presented in Section 6.5.

In reporting these floor dust lead results, the researchers are cognizant that this report only provides information on one element of the decision making process that goes into selecting standards. Practical issues of implementation such as the reliability of low-cost dust lead laboratory analysis at low (<25 µg/sample) levels would need to be addressed.

The analysis of this objective is also limited due to the fact that the study populations in the two pre-1950 sites were so different. While the original study design envisioned a study size of 150 pre-1950 dwellings, the lack of variation between environmental lead levels in homes with and without an EBL child in New York City caused the researchers to focus on the 64 dwellings from Milwaukee. Additional studies at other sites could help clarify the interpretation of the findings presented in this report.

6.1.3 Objective 3c: What should the maximum single surface standards be, assuming HUD's standards apply to averages? What should the average standards be, assuming HUD's standards apply to maximum of single surface samples?

6.1.3.1 Methodology

During the federal regulatory process that was initiated by Title X, HUD and EPA have used empirical data to determine a single dust lead standard for each surface type (floors, window sills and window troughs). Of interest is that the same dust lead standard is used whether the standard is applied to the household maximum (as is used for single-surface sampling) or the household mean (as is used for composite sampling). Because the laws of mathematics dictate that a home is more likely to fail when using the same standard for the maximum as for the mean, a certain level of inequity has been built into the system. A property owner who seeks to reduce the chance of failing a risk assessment would be foolish not to demand that composite sampling be used. Only EPA's Section 403 rules cope with this inequity by requiring that a weighted-mean be applied to all sampling protocols and denying the use of household maximums. Yet, even with this rule, a household maximum is used for clearance purposes.

Analyses were conducted to determine the appropriate levels needed to create equivalent dust lead standards. The first consideration was whether the current standards are based on the predictive power of a mean or maximum dust lead level. Although much of the empirical research that was used in the regulatory process established relationships between average dust lead loadings and blood, the current standards were ultimately determined after combining this research with feasibility issues [4]. Therefore, current standards cannot be considered either mean-based or maximum-based. Two separate analyses were therefore conducted, using the alternative positions that the current standards are mean-based or maximum-based.

A bivariate regression model was used to predict the log of maximum dust lead loadings based on the log of the mean dust lead loadings. The R^2 values were 0.97, 0.98 and 0.99, for floors, window sills and troughs, respectively. This regression equation was then used to estimate the maximum lead loading that corresponds to a mean lead loading for each surface type. The same method was used to find the mean lead loading that is equivalent to a maximum. The analyses were conducted on all eligible dwellings in Milwaukee and New York City that had at least three samples per surface type per dwelling. Analyses were conducted for floors (uncarpeted, carpeted, all), window sills and window troughs. Final results are presented for each of these surfaces with the exception of carpeted floors. The mean and maximum dust lead results on carpeted floors in the 19 dwellings that met the inclusion criteria all fell below the current standards. The results are presented for all of the current standards for risk assessments and lead hazard screens as well as for a floor dust loading of 15 $\mu g/ft^2$ (representing a lower more optimal standard). These analyses did not consider blood lead levels and are solely intended to identify equivalent standards without regard to predictive power.

6.1.3.2 Findings

The results of the analyses of equivalent dust lead standards are presented on Table 6.1.6. The third column presents the results of what an equivalent *maximum* dust lead standard should be if the current standards are mean-based. The fourth column presents the results of what an equivalent *mean* dust lead standard should be if the current standards are maximum-based. For example, a dwelling that HUD considers to have a dust lead hazard when a single all floor dust sample is at or above 40 $\mu g/ft^2$, should be considered hazardous if the composite (mean) dust sample is at or above 26 $\mu g/ft^2$.

6.1.3.3 Discussion of Findings

The findings demonstrate that there are clear and fairly substantial differences between the standards that would be appropriate for a single-surface maximum household dust lead measurement and those appropriate for a composite mean dust lead measurement. The results suggest that a maximum standard should be at least 50 percent higher than a mean standard. If both mean and maximum criteria for failure are allowable under risk assessment protocols, separate standards by sample type should be considered.

If such separate standards are considered during a regulatory process, the agencies should keep a number of factors in mind. They must determine whether the empirical studies that are the basis for standard setting used means or maximum dust lead levels to establish relationships with blood lead. They should also consider how many samples

Table 6.1.6: Equivalent Dust Lead Loading Standards
Using 3 Samples per Sample Type per Dwelling

Sample Type	Dust Lead Loading Standard (μg/ft^2)	Equivalent Max Standard assuming Current Standard based on Mean (μg/ft^2)	Equivalent Mean Standard assuming Current Standard based on Maximum (μg/ft^2)
Uncarpeted Floor N=97	15	22	10
	25	38	16
	40	62	25
	50	79	31
	100	163	59
All Floors N=147	15	22	10
	25	37	17
	40	61	26
	50	77	32
	100	159	60
Window Sills N=111	125	223	73
	250	460	140
	500	951	268
Window Troughs N=93	400	716	232
	800	1,453	454

would be collected on a particular surface under the recommended sampling plan, because the relationship between the mean and the maximum will vary by sample size. (The relationship identified in this analysis was based on a sample size of three samples per surface.) Finally, the regulators should consider the variability of dust lead levels within a dwelling. As discussed further in Section 6.4, the variability does appear to vary by site and possible age of housing, which would affect the recommended differences in standards.

Finally, it must be emphasized that these analyses set out to consider differences between the mean and maximum measures of dust lead without regard to the predictive power of either of these measures. Just because these analyses suggest that it would be appropriate to set the composite (three-sample) all floor dust lead standard at 26 μg/ft^2 when the single-surface maximum standard is 40 μg/ft^2 does not mean to imply that either of these standards is a valid predictor of the blood lead outcomes (EBL/non-EBL).

6.1.4 Objective 3f: Compare true composite samples and mathematically averaged single surface samples taken in the same locations.

6.1.4.1 Methodology

The analyses of the risk assessment protocols in Section 5 and the analyses up to this point in Section 6.1 used the arithmetic mean of single-surface samples when considering protocols that call for the standard to be applied to composite samples. This methodology was used because it gave the study more flexibility when considering mean values. While the number and location of the subsamples in the composite samples collected by the risk assessors are fixed, these factors could be manipulated as needed when analyzing the mean of the single-surface samples. However, this approach cannot be justified unless it can be proven that the average of single-surface samples is equivalent to actual field composites.

To investigate this hypothesis, composite samples and mathematically averaged single surface samples taken in the same locations were compared. Data were only used when a perfect match was possible and the number of sub-samples in the composite was between 2 and 4 (inclusive). Enough data were available for uncarpeted floors (central location), uncarpeted floors (under window), window sills and window troughs to be included in the models. Less than ten dwellings had composite samples collected from carpeted floors where the single surface dust data matched perfectly and were not included in the analyses. The data that was included in the analyses is presented on Table 6.1.7.

Table 6.1.7: Number of Dwellings included in Model by Sample Type and Site

Surface Type	Data in Model	Number of Dwellings per Sample Type & Site	Total Number of Dwellings
Uncarpeted Floors (Central)	Milwaukee 2-sample	21	65
	Milwaukee 4-sample	14	
	New York 4-sample	30	
Uncarpeted Floors (Under Window)	Baltimore Co 2-sample	22	95
	Milwaukee 2-sample	17	
	Milwaukee 4-sample	11	
	New York 2-sample	10	
	New York 3-sample	18	
	New York 4-sample	17	
Window Sills	Baltimore Co 4-sample	77	169
	Milwaukee 4-sample	43	
	New York 2-sample	14	
	New York 3-sample	16	
	New York 4-sample	19	
Window Troughs	Baltimore Co 4-sample	70	156
	Milwaukee 4-sample	38	
	New York 2-sample	20	
	New York 3-sample	12	
	New York 4-sample	16	

A mixed model was run to explore the effect of sample type (composite or mean single surface) on either the geometric mean dust lead loading results or the variability of those results. The researchers were concerned that should either geometric mean dust lead

loadings or their variability differ by collection method, it would not be possible to infer anything about composite sampling from previous analyses that used mean dust lead levels. Beyond this study, a lack of relationship between the two collection methods would call into question any standard setting for composite samples that does not use composite sampling as its scientific basis.
In the mixed model, the effect of the four variables below on the log dust lead loading was tested, as was the effect of the first three variables on the variability of the dust lead.

1. Site (Baltimore County, Milwaukee or New York City)
2. Type of sample (composite or the average of single surface)
3. The number of sub-samples (2 to 4)
4. The interaction of type of sample and number of sub-samples

See Appendix C.1 for technical details.

6.1.4.2 Findings

With the exception of window sills, the results were generally consistent: after controlling for site and number of samples collected, dust lead loading and its variability did not depend on whether the sample was a composite sample or the average of single surface samples. The modeling of uncarpeted floors and windows troughs found geometric mean dust lead loadings were significantly different by site ($p<0.001$ for each), and their variability was different by site and number of samples collected. Further discussion about the variability with respect to number of samples collected will be presented in Section 6.4.

Window sills samples acted very differently. The geometric mean dust lead loading of composite samples were significantly different from the geometric mean of the mean single surface samples ($p<0.001$). The composite dust lead loadings were estimated to be 33 percent lower than the dust lead loadings for arithmetically averaged sills. The geometric mean dust lead loading results were also significantly different by site ($p<0.001$), but none of the variables in the model were significantly related to the variability of results.

6.1.4.3 Discussion of Findings

The results are generally supportive of the decision to use the average of the single-surface samples as a surrogate for composite samples in this report. One exception was window sill samples. They did perform differently raising some concerns about the equality of mean single surface window sill results and composite sample results. This result might suggest that window sill loadings in analyses using mean dust lead loadings should be discounted by 33% when they are to represent composite sampling results. Presumably, this should be done because laboratory recovery rates are lower for composite samples than single surface samples. However, the difference between single sample and composite quality control sample recovery rates (98% v. 93%) did not identify a bias even approaching -33%. Therefore, there is not persuasive evidence to consider adjusting the mean dust lead level for any surface type when considering the effect of composite sampling.

The lack of sample size for carpeted composite prevented the study from considering the equivalency of the two collection methods on carpeted surfaces. Further study specifically designed to collect matching composite and single surface carpeted floor samples is needed to confirm that the collection methods are equivalent on carpets.

6.2 Investigation of the Sources of Dust Lead Loading including Friction/Impact, Blow-in and Track-in

6.2.1 Purpose

As defined in Title X, two of the six lead-based paint hazard conditions are lead-based paint on friction surface and lead-based paint on an impact surface. Lawmakers were concerned that leaded components that were subject to frequent abrasion or impact would be active sources of leaded dust. While common logic supports the legislators' position, little scientific study has been conducted to assess the relative magnitude of leaded dust that is generated from these potential sources.

Beyond the issue of the significance of friction/impact surfaces as sources of dust lead, lies the practical problem of including a test for friction/impact hazards into a risk assessment protocol. The HUD *Guidelines* stated that, "Operating three of four windows and three or four doors is usually adequate; it is not necessary to operate all windows and doors in a dwelling. For risk assessment purposes, it is not necessary to analyze the paint for lead content on these surfaces unless it is deteriorating." Practically, however, this guidance has not been incorporated into a risk assessor's determination of whether a dwelling fails. A risk assessor would be hard pressed to state that a dwelling with sticking doors but no other hazards must go through the process of treatment, clearance and re-evaluation.

In its final Section 1012/1013 rules, HUD recognized that unless there is another lead-based paint hazard present, a friction/impact surface should not be considered a hazard. Friction surfaces must be treated if the dust lead loadings on an adjacent component are above the standard *and* there is evidence of abrasion and the surface is leaded or presumed to be leaded. Impact surfaces are to be treated if there is deteriorated paint that is known or presumed to be leaded *and* there is evidence of impact from a building component. In essence, this rule does not add any factors to the risk assessment protocols in order to determine risk. It simply adds guidance to help the risk assessor develop a hazard control plan.

While the primary purpose of this report is to evaluate the ability of a risk assessment to determine risk, this particular objective is directed toward the development of a hazard control plan. The study examines whether there is evidence that friction surfaces were significantly related to dust lead loadings, which would justify requiring lead hazard control activities. Furthermore, if a relationship was found, consideration must be given to whether the current dust sampling protocols adequately identify the potential hazard. Because the study design only called for the risk assessors to evaluate whether windows and doors were rubbing or binding, an assessment of the influence of impact is outside the scope of this study. However, because it is necessary to control for all pathways of lead into dust in order to evaluate the influence of friction, the effect of interior and exterior sources of lead (e.g., blow-in/track-in) were examined.

6.2.2 Objective 4a: Investigate the sources of dust lead loading including friction, impact, blow-in and track-in to uncarpeted floor dust lead loading.

6.2.2.1 Methodology

As a first analysis, the study considered the relationship between friction on windows and doors and floor dust lead loadings. A mixed model was used, to account for possible correlations between measurements within the same household and room. A backward regression procedure with forward steps as needed was followed. The log dust lead loading for any available floor sampling location in a room was used as the dependent variable, while the risk assessor's assessment of rubbing or binding and the paint lead levels on the corresponding window were used as the variables of interest. All other possible sources of lead that might contribute to the floor dust lead loadings were considered including paint lead, soil lead, and factors related to their potential contribution, such as soil cover, dwelling height and accessibility of any windows.

The possible pathways of lead that are accounted for in the model included:

1. Window friction, window paint condition and window paint lead
2. Door friction, door paint condition and door paint lead
3. Lead paint (and condition) of the room
4. Exterior Lead Sources (Soil lead, other point sources)
5. Blow-in from the exterior
6. Track-in from the exterior

In addition to these variables, a set of general variables (site, building age, location of wiped surface, condition of wiped surface, etc.) was included to control for possible confounding factors. Appropriate interactions of variables were also included. A complete list of variables with explanatory details is presented in Appendix C.2.

Because the results were unstable when rooms with missing data were allowed in the analysis dataset, only rooms with no missing variables were used. Complete data were available for 782 floors in 209 rooms in 173 dwellings.

6.2.2.2 Findings

Variables determined to be significant predictors of uncarpeted floor dust lead loading were:

- Interaction of site, window height and exterior dust lead loading (represents blow -in from exterior).
- Sample location on floor (window, central, entry or perimeter)
- Interaction of site and soil lead concentration. (represents blow-in/track-in)
- Interaction of location of floor and condition of window trough paint
- Average window paint lead loading .
- Interaction of average door paint lead loading and floor location
- Outside hall/porch dust lead loading

When interactions are listed, all variables that are part of the interaction were also included in the model.

The model was determined to be heteroscedastic, that is, the variability of the building effects and room effects depended on the site.

6.2.2.2 Discussion of Findings

The variables of interest, the interaction between the observation of rubbing/binding on doors and door paint lead and the interaction between the observation of rubbing/binding on windows and window paint lead, were not significantly related to the floor dust lead loadings in this model. Assuming that window or door friction does produce dust lead, these results indicate that floor sampling would not be a good measure of its occurrence.

While of less practical importance in terms of a risk assessment, the statistical modeling supports the position that lead-contaminated dust comes from multiple sources through multiple pathways. In addition to the principal interior source of lead, lead-based paint, the other significant variables in the model support the position that lead is transported into a dwelling through blow-in and track-in. Two exterior sources, soil lead and exterior dust lead, are identified in the model, but the variability of the findings by site also suggests that the differing levels of ambient street lead and air lead may also be sources.

This analysis used a mixed model so that multiple observations within buildings and rooms could be appropriately accounted for. Unfortunately, a determination of measures such as percentage of variation accounted for in the model and percentage of variation accounted for by a variable is not possible with the mixed model. However, while such knowledge would be useful to understanding the pathways of lead, it is outside of the principal objective of this study, which is to understand how risk assessment protocols can become better predictors of risk.

6.2.3 Objective 4b: Investigate the sources of dust lead loading including friction, impact and blow-in to window sill dust lead loading

6.2.3.1 Methodology

The relationship between window friction and the dust lead loadings on window sills was assessed using the same analytical approach that was used to assess the influence of friction on floor dust lead loading. With exception of a reduction in variables, the same mixed modeling analysis was used employing a backward regression procedure followed by forward steps as needed. Because door friction was not considered a likely source of leaded dust on window sills, the variables representing it were not included in the model. Likewise, the variables representing track-in of exterior leaded dust were also dropped from the list of possible confounding factors.

Using the window paint condition and rubbing/binding variables, a new window variable with 4 levels was created:

- Intact paint, no rubbing/binding
- Non-intact paint, no rubbing/binding
- Intact paint, rubbing/binding
- Non-intact paint, rubbing/binding

The possible pathways of lead accounted for in the model included:

1. Window rubbing/binding, window paint condition and window paint lead
2. Lead paint (and condition) of the room
3. Exterior lead sources (soil lead, other point sources)
4. Blow-in from the exterior

In addition to these variables, a set of general variables (site, building age, condition of wiped surface, etc.) was included to control for possible confounding factors. A complete list of variables with explanatory details is presented in Appendix C.2.

Because the results were unstable when rooms with missing data were allowed in the analysis dataset, only rooms with no missing variables were used. Complete data were available for 611 window sills in 182 dwellings.

6.2.3.2 Findings

The occurrences following window conditions were observed: (1) 44% intact paint with no rubbing/binding; (2) 18% intact paint with rubbing/binding; (3) 8% some non-intact paint with no rubbing/binding; and (4) 30% some non-intact paint with rubbing/binding. The frequency of rubbing/binding was much higher for windows with non-intact paint than for windows without non-intact paint. Overall 48% of the windows had rubbing/binding, but 30 percent of the windows with intact paint had rubbing/binding versus 79 percent of windows with non-intact paint.

Variables determined to be significant predictors of window sill dust lead loading were:

- Cleanability of wiped window sill
- Site
- Interaction of window paint condition and rubbing/binding
- Interaction of average interior window paint lead, window average paint condition and rubbing/binding. (represents window lead hazards)
- Condition of window trough paint
- Exterior dust lead loading

When interactions are listed, all variables that are part of the interaction were also included in the model.

Windows with rubbing/binding, non-intact paint or both have the same windowsill loading (across the three quartiles of window paint loading p=0.258). Windows with rubbing/binding and/or non-intact paint have higher windowsill loadings than windows with neither of these conditions (across the three quartiles of window paint loading (0.1 (1st and 2nd) and 2.5 mg/cm^2 (3rd); $p<0.001$). At a window paint loading of 1 mg/cm^2, windows with intact paint and no rubbing/binding are expected to have windowsill dust loadings that are 27% lower than windows with non-intact paint and/or rubbing/binding.

6.2.3.3 Discussion of Findings

The analysis that examined the relationship between window friction and window sill dust lead loadings supports the conclusion that window friction is a significant source of window sill dust lead even when window paint is intact. Dust lead loadings were higher on window sills where rubbing or binding was identified or window paint was non-intact and dust lead loadings on those windows increased with the levels of paint lead. If no non-intact paint and no rubbing/binding were observed, window paint lead levels were not associated with window sill dust lead levels.

The results support the current practice of using dust testing to identify windows that may be candidates for fiction or impact controls. Evidence of non-intact paint is not adequate to determine where friction/impact problems may exist. If sampling is not conducted on a window sill, then the window may need to be opened to access rubbing or binding and determine if treatment is needed.

6.3 Assessment of the Ability of the Current HUD Paint Film Classification System to Predict Dust Lead Loading & Elevated Blood Lead

6.3.1 Purpose

Chapter 5 of the HUD *Guidelines* included a table to help risk assessors judge the level of deterioration on a painted surface. The paint assessment system, known as the HUD Paint Film Classification System, created a standardized process for describing the extent of lead-based paint deterioration. The system divides the paint film into three categories: intact, fair and poor (Table 6.3.1).

Table 6.3.1: Categories of Paint Film Quality (HUD Guidelines Table 5.3)

	Total Area of Deteriorated Paint on Each Component		
Type of Building Component	Intact	Fair	Poor
Exterior components with large surface areas.	Entire surface is intact	Less than or equal to 10 square feet	More than 10 square feet
Interior components with large surface areas (walls, ceilings, floors, doors)	Entire surface is intact	Less than or equal to 2 square feet	More than 2 square feet
Interior and exterior components with small surface areas (window sills, baseboards, soffits, trim)	Entire surface is intact	Less than or equal to 10 percent of the total surface area of the component	More than 10 percent of the total surface area of the component

The HUD *Guidelines* state that paint in poor condition is a lead-based paint hazard, while paint in fair condition should be monitored. In 1999, however, HUD broke from its previous guidance and determined that all deteriorated paint, no matter how minimal, should be considered a hazard. The final HUD Section 1012/1013 rule, as well as the final EPA Section 403 rule, considers all non-intact leaded paint to be a lead-based paint hazard

Although HUD no longer recognizes its Paint Film Classification System, the study proceeded with its original analysis plan to consider how well paint condition as defined by the HUD System correlates with dust lead loading and blood lead outcomes and if improvements could be made. The study has also examined how well inspectors could implement the system (i.e., it’s ease of replication).

6.3.2 Objective 5a: Does the presence of lead-based paint in “poor” condition significantly contribute to the prediction of uncarpeted floor dust lead loading?

6.3.2.1 Methodology

A mixed model was used to assess the relationship between the condition of leaded paint and average dust lead loading on uncarpeted floors in a room. In this model, influences common to the dwelling are accounted for by the inclusion of a random dwelling effect.

A mixed model is an appropriate method based on the assumption that any differences in dust lead loadings between rooms are dependent on the most deteriorated leaded surface and error. Risk assessors determined the condition of the paint using the HUD Paint Film Classification System (Intact, Fair, Poor). In the model, the dependent variable was the log average of the four sampling locations on the floor within the room (i.e., entry, central, perimeter and under window). Three variables were included in the model:

- Site (Milwaukee, New York, Baltimore County)
- Condition of the most deteriorated component with lead-based paint in the room (intact, fair, or poor)
- Interaction of site with condition of most deteriorated component with lead-based paint

Both main effects and the interaction term were significant in this model; therefore separate regression analyses were used for each site with the paint condition variable retained as the only independent variable. When a statistically significant relationship was identified between paint condition and floor dust lead loading, two hypotheses were tested: 1) Fair is equivalent to Poor, and 2) Good is equivalent to Fair. If either hypothesis were accepted, the finding would support combining categories. The model also estimated the incremental change of the uncarpeted floor dust lead loading by paint category. Results were considered significant if $p<0.05$.

6.3.2.2 Findings

The results varied by site and offered no consistent pattern. In Baltimore County and New York City, the most deteriorated leaded component in a room was not a significant predictor of floor dust lead loading, indicating that there is no reason to consider collapsing variables. In Milwaukee, the most deteriorated leaded component was a significant predictor of floor dust lead loading. At that site, Fair and Poor condition categories could be combined but not Good and Fair.

Table 6.3.2: Relationship between Lead-Based Paint Condition and Dust Lead Loading by Site

Site	Number of Rooms	Relationship between LBP Condition and Dust Lead Loading	Equivalence between Good and Fair Condition	Equivalence between Fair and Poor Condition
Baltimore County	296	No P=0.68	N/A	N/A
Milwaukee	223	Yes P<0.01	No P<0.01	Yes P=0.26
New York City	237	No P=0.20	N/A	N/A

N/A=Not applicable

Using these results, Fair and Poor conditions were collapsed into one category in Milwaukee and the model was rerun. The most deteriorated leaded component remained a significant predictor of uncarpeted floor dust lead loading ($p<0.01$). Rooms where the worst paint condition was fair/poor were expected to have a dust lead loading 92% higher than those rooms with only intact leaded paint (95% confidence interval: 43% higher to 157% higher).

6.3.2.3 Discussion of Findings

The condition of the most deteriorated leaded paint in a room (without regard to the specific paint lead level) was an inconsistent predictor of the uncarpeted floor dust lead loadings in that room. In the dwellings in Baltimore County and New York City, a relationship between the paint in worst condition and the floor dust lead loading could not be established. In the dwellings in Milwaukee, a relationship did exist which supported the hypothesis that deteriorated leaded paint is contributing to the dwelling's floor dust. The results of these analyses suggest that in communities that have similar housing conditions as Milwaukee, it may be appropriate to define non-intact leaded paint as a hazard. This issue will be further considered in Sections 6.3.5 and 6.3.6.

6.3.3 Objective 5b: If the definition of "poor" were modified, would the ability of the paint film classification system to predict uncarpeted floor dust lead loading improve?

6.3.3.1 Methodology

Like the previous models, a mixed model was used to assess the relationship between the condition of leaded paint and average dust lead loading on uncarpeted floors. Instead of using the three-category HUD Paint Film Classification System, a 6-level paint condition scale was applied (Table 6.3.3). Once again, the dependent variable was the log average of the four sampling locations on the floor in the room (entry, central, perimeter and under window). Based on the findings from the previous analysis, a separate model was run for each site. The independent variable was the condition of most deteriorated component with lead-based paint in the room.

When an analysis identified a significant relationship between the paint condition and the floor dust lead loading the study considered the five possible ways that the six levels of paint condition could be collapsed into two categories (low deterioration/high deterioration) of paint condition. The cut points between the six levels defined the categories (Table 6.3.4). The equality of the geometric mean floor dust lead loadings within the low deterioration and within the high deterioration categories was tested for each of the five splits. Equality within each of the two deterioration categories indicated that the levels of paint condition within each deterioration category could be collapsed. When combined with the earlier finding that paint condition was related to dust lead loadings, such an outcome would identify a difference in effects on floor dust lead loadings between the low deterioration category and the high deterioration category. Results were considered significant if $p<0.05$.

Table 6.3.3 NCHH Study Categories of Paint Deterioration

Type of Building Component	Level of Paint Film Deterioration					
	1	2	3	4	5	6
Exterior components with large surface areas	Intact	Less than or equal to 5 square feet	More than 5 ft^2 and less than or equal to 10 ft^2	More than 10ft^2 and less than or equal to 15ft^2	More than 15ft^2 and less than or equal to 25 ft^2	More than 25 square feet
Interior components with large surface areas (walls, ceilings, floors, doors)	Intact	Less than or equal to 1 square foot	More than 1 ft^2 and less than or equal to 2 ft^2	More than 2 ft^2 and less than or equal to 3 ft^2	More than 3 ft^2 and less than or equal to 10 ft^2	More than 10 square feet
Interior and exterior components with small surface areas (window sills, baseboards, soffits, trim)	Intact	Less than or equal to 5% of the total surface area	More than 5% and less than or equal to 10% of surface area	More than 10% and less than or equal to 15% of surface area	More than 15% and less than or equal to 25%	More than 25% of the total surface area

6.3.3.2 Findings

In Baltimore County and New York, the most deteriorated component with lead-based paint (using the 6-level scale) in a room was not significantly related to uncarpeted floor dust lead (p=0.77 and p=0.49, respectively). In Milwaukee, the most deteriorated LBP component using the modified scale was a significant predictor of floor dust lead loading (p<0.01). Because paint condition using the 6-level scale was a significant predictor of floor dust lead loading only in Milwaukee, the effects of the different measures of high and low deterioration were only tested at that site (Table 6.3.4).

Table 6.3.4: Comparison of Paint Deterioration Categories (Milwaukee)

Low Deterioration Category	High Deterioration Category	Equality within Low Deterioration and High Deterioration Categories	Increase in GM Floor Dust Lead Loading from Low to High Det. (95% Conf. Int.)
Level* 1	Levels 2-6	Yes P=0.655	52% (39%, 70%)
Levels 1-2	Levels 3-6	Yes P=0.115	62% (47%, 80%)
Levels 1-3	Levels 4-6	No P=0.010	N/A
Levels 1-4	Levels 5-6	No P=0.011	N/A
Levels 1-5	Level 6	No P=0.012	N/A

*Levels defined on Table 6.3.3

Of the five groups of low/high deterioration categories examined, equality within the two deterioration categories was found for two of the groups: (1) the group in which the low deterioration category combined Levels 1-2 and the high deterioration category combined Levels 3-6(p=0.115); and (2) the group in which the low deterioration category combined Levels 1 and the high deterioration category combined Levels 2-6 (p=0.655). Although these results indicate that level 2 could be combined with level 1 or with 3-6, the p-value gives stronger evidence that level 2 should be combined with level 3-6 than levels 2 (0.655>0.115).

6.3.3.3 Discussion of Findings

The 6-level scale for paint deterioration was created to try to classify higher levels of deterioration because of the possibility that only the most seriously deteriorated surfaces would contribute significantly to dust lead loadings. Like the analysis of the HUD Paint Film Classification System, the 6-level paint condition scale performed inconsistently across sites. The level of paint deterioration on components with lead-based paint was a predictor of floor dust lead loading only in Milwaukee. In Milwaukee, the best cut-point for lead-based paint was when poor lead-based paint was defined as non-intact paint, indicating that the additional detail of the 6-level deterioration coding does not add any predictive strength. The results in this section are consistent with the results presented in Section 6.3.2.

6.3.4 Objective 5c: Is the current paint film classification system replicable by two risk assessors?

6.3.4.1 Methodology

A pair of risk assessors conducted each risk assessment. As part of the study design, both risk assessors were expected to assess the paint condition in 30 dwellings at each site. These reliability assessments were conducted separately so that one risk assessor was blind to the second one’s findings. Table 6.3.5 presents the number of dwellings, rooms and components where two risk assessors visually assessed the paint condition.

Table 6.3.5: Dwellings, Rooms and Components Tested by both Inspectors

Site	Inspectors	Number of Dwellings	Number of Rooms	Number of Components
Baltimore County	#1 & #2	6	92	673
	#1 & #3	3	47	287
	#1 & #4	3	36	242
	#2 & #4	9	129	962
	#3 & #4	15	201	1304
	All	36	505	3468
Milwaukee	#5 & #6	17	225	1622
New York	#8 & #9	3	33	163
	#8 & #10	30	349	2159
	All	33	382	2322
All	All	86	1112	7412

Eight different pairs of risk assessors conducted these reliability assessments for this study: five in Baltimore County, one in Milwaukee and two in New York City. Each pair conducted reliability assessments in 3 to 30 dwellings. Because each risk assessor was blind to the other's work, and the study did not encourage any follow-up review, the number of dwellings assessed together would not be expected to impact the findings. Each pair assessed at least 163 components, providing a large enough sample to examine all eight pairs.

Using these data, the study examined the replicability of the HUD Paint Film Classification System by risk assessors in the field. The study presented two measures that test the concordance of the findings for each pair of risk assessors. The first was the inter-rater reliability, which was reported using the weighted-Kappa statistic, a measure similar to the correlation coefficient except that it is used for ordinal data. As the measure approaches 1.0, a stronger correlation of results is demonstrated. The second measure was the test of equality of mean rating. In this analysis, the test of equality was used to see if one rater tended to give better condition ratings than the other. The study also presented reliability results for the intact/non-intact condition coding system now dictated by HUD.

6.3.4.2 Findings

The inter-rater reliability for the eight pairs of risk assessors using the HUD Paint Film Classification System (Intact, Fair, Poor) ranged from 0.922 to 0.438 (Table 6.3.6). The mean inter-rater reliability was 0.667. The test of equality found that the mean rating for the two risk assessors was significantly different for five of the eight pairs. In two cases, however, the magnitude of the difference in the mean rating was less than 0.05, but the assessors examined so many components together (over 1,000 in each case), that the difference was significant. In general, the risk assessors reached similar conclusions about the condition of paint, but the consistency of the results were lower than what might be considered ideal.

When the 2-level condition coding system was considered (Table 6.3.7), the results were similar to the 3-level system. The inter-rater reliability ranges from 0.915 to 0.459. The mean inter-rater reliability was 0.670. For each individual pair, the results were within the confidence interval for the 3-level condition coding system. The results were also very similar for the test of equality. Five of the eight pairs had statistically significant differences between the mean condition ratings, but in two cases those differences were less than or equal to 0.05.

Table 6.3.6: Rater Concordance for the HUD Paint Film Classification System (3 Levels: Intact, Fair, Poor)

Site	Risk Assessors	Inter-rater Reliability (95% CI)	Mean Rating[1] for 1st Inspector	Mean Rating[1] for 2nd Inspector	P-value for test of equality of mean rating
Baltimore County	#1 & #2	.922 (.891, .993)	1.27	1.28	0.221
	#1 & #3	.506 (.414, .598)	1.33	1.42	0.002*
	#4 & #1	.574 (.482, .665)	1.48	1.46	0.697
	#4 & #2	.813 (.767, .858)	1.21	1.24	0.002*
	#3 & #4	.730 (.689, .770)	1.28	1.28	0.811
	All	.759 (.735, .782)			
Milwaukee	#5 & #6	.652 (.619, .685)	1.42	1.56	<0.001*
New York	#8 & #9	.438 (.314, .562)	1.57	1.40	<0.001*
	#8 & #10	.702 (.674, .730)	1.48	1.52	<0.001*
	All	.685 (.657, .713)			
All	All	.714 (.699, .730)			

[1] Based on ranking 1=intact, 2=fair, and 3=poor.

Table 6.3.7: Rater Concordance for the 2-Level Paint Condition Code (Intact=Intact, Not Intact=Fair & Poor)

Site	Inspectors	Inter-rater Reliability (95% CI)	Mean Rating[1] for 1st Inspector	Mean Rating[1] for 2nd Inspector	P-value for test of equality of mean rating
Baltimore County	#1 & #2	.915 (.876, .954)	1.19	1.20	0.346
	#1 & #3	.535 (.434, .635)	1.30	1.39	0.001*
	#4 & #1	.611 (.510, .712)	1.43	1.43	1.000
	#4 & #2	.801 (.751, .851)	1.17	1.19	0.001*
	#3 & #4	.724 (.681, .768)	1.25	1.25	1.000
	All	.750 (.724, .776)			
Milwaukee	#5 & #6	.630 (.593, .668)	1.32	1.45	<0.001*
New York	#8 & #9	.459 (.331, .588)	1.55	1.40	<0.001*
	#8 & #10	.682 (.651, .713)	1.42	1.47	<0.001*
	All	.665 (.635, .696)			
All	All	.702 (.685, .719)			

[1] Based on ranking 1=intact, 2=not intact (fair & poor).

6.3.4.3 Discussion of Findings

Risk assessors demonstrated that the HUD Paint Film Classification System could be reliably used to assess the condition of paint in the field. The pairs of inspectors concurred at a rate of 67 percent. While this level of concordance may appear low for a simple test, the level is exactly the same as for the most basic test of deterioration (intact/non-intact). Although critics have stated that the HUD 3-level System is complicated and difficult to use, it performed as well as the simpler 2-level condition coding system. These results indicated that concerns about field implementation should not be a factor when determining the best method to identify deteriorated lead-based paint.

6.3.5 Objective 5d: If the definition of "poor" were modified would the ability of the paint film classification system to predict EBL children improve?

6.3.5.1 Methodology

Prior to running the final analyses to address this objective, a logistic regression analysis was conducted to assess the impact of site on the relationship between the condition of leaded paint using the HUD Paint Film Classification System (Intact, Fair, Poor) and the blood lead outcome for a home. In the model, the dependent variable was the presence of an EBL child (yes/no) at the dwelling. Three variables were tested in the model:

- Site (Milwaukee, New York, Baltimore County)
- Condition of most deteriorated interior LBP painted component in the dwelling (intact, fair, or poor)
- Interaction of site with condition of most deteriorated LBP component

Because Baltimore County had only one child with an elevated blood lead level, the model would not converge and hence, could not offer any results. The model was then rerun without Baltimore County data. All variables were significant in the revised model, indicating that separate logistic analyses should be conducted for Milwaukee and New York City.

Four logistic analyses were conducted, one for each combination of site (Milwaukee and New York) and paint hazard definition ("poor" under the former HUD System or non-intact). The only independent variable included in the model was presence/absence of a paint hazard as determined by the condition of the most deteriorated leaded component in the dwelling. When a relationship between condition and blood lead outcome was established, the model reported the odds-ratio or incremental change in the probability of finding a child with an elevated blood lead level by paint category. Results were considered significant if $p<0.05$.

6.3.5.2 Findings

In both Milwaukee and New York, the presence of a single *non-intact* leaded interior component was not a significant predictor of a home with a child with an elevated blood lead level (p-values of 0.09 and 0.91, respectively). In Milwaukee, the presence of a single interior leaded component in *poor* condition was predictive of a home with an EBL child (p=0.02). In this study, a dwelling in Milwaukee with any *poor* leaded paint on the interior was 29.1 times more likely (95% CI: 2.4 to 113.4 times more likely) to house a child with an elevated blood lead level than a dwelling with leaded paint that was intact or in fair condition. In New York, the presence of a single leaded component in *poor* condition was a significant predictor (p=0.04), but of a dwelling with a child ***without*** an elevated blood lead level.

6.3.5.3 Discussion of Findings

The results suggest that the presence of a single leaded component in deteriorated condition is, at best, an inconsistent indicator of risk in pre-1950, urban dwellings. Only in Milwaukee was a significant relationship demonstrated between paint condition and the presence of an EBL child, and only when deteriorated paint was defined as “poor.” In New York City, the finding of a correlation in the “wrong” direction must be considered spurious. This result points out how spurious relationships can be identified when examining a hazard with multiple sources. Obviously, there must be other factors that are putting children at risk in the New York City dwellings that do not have interior leaded paint in “poor” condition.

The results for Milwaukee correspond to the findings in Section 6.3.2 on this report, which examined the relationship of paint lead condition to floor dust lead. The presence of non-intact interior leaded paint was neither predictive of dust lead loadings nor blood lead outcomes, while the presence of interior leaded paint in “poor” condition was predictive of both outcomes. For New York City, the results are complicated by the spurious finding in this analysis, but findings also indicate that non-intact interior leaded paint was neither predictive of dust lead loadings nor blood lead outcomes at this site.

6.3.6 Objective 5e: What is the best paint lead measure for predicting EBL?

6.3.6.1 Methodology

In each of the previous sections concerning Objective #5 where a relationship between paint condition and blood or dust lead loading was examined, the analysis looked at the single most deteriorated leaded component in the room or dwelling. An alternative hypothesis is that a better predictor of risk would not be a single deteriorated painted surface, but a certain number of deteriorated surfaces. For example, one might hypothesize that a dwelling should not be considered “high-risk” property unless five leaded components were identified as deteriorated.

Three factors were considered in the analyses:

- Is deterioration of any paint or just lead-based paint a better predictor of risk?
- Is non-intact paint or “poor” paint condition a better predictor of risk?
- What frequency of deterioration is a better predictor of risk (1, 2, 5, 10, 20, 30)?

The six frequencies selected for the third factor were somewhat arbitrarily picked after examining the data on the frequency of reported deteriorated surfaces.

Based on the limitations with the data described in section 6.3.5.1, Baltimore County data were excluded from the analyses and separate analyses were conducted for Milwaukee and New York City. Each combination of site and paint variable (Any LBP, Non-Intact LBP, Poor LBP, Non-Intact Paint, and Poor Paint) was considered. For each of the five paint variables, the six different frequencies of paint in the category of interest were considered as categorical variables resulting in 30 measures of paint condition at each site. For example, for the test to investigate the predictive power of any lead-based paint, the effect on the likelihood of a dwelling housing an EBL child when the dwelling had 1, 2, 5, 10, 20 or 30 surfaces with lead-based paint was assessed. Using Fisher’s exact test of independence, the relationship between elevated blood lead (yes/no) and the paint measure was assessed.

6.3.6.2 Findings

One measure of lead-based paint condition, the number of surfaces with lead based paint in *poor* condition, was a significant predictor of the blood lead outcome (EBL/no EBL) in Milwaukee (Table 6.3.8). In Milwaukee, the only number of surfaces that were predictive of a high-risk dwelling was one surface in poor condition. A dwelling with at least one surface with poor LBP was 126% more likely to house an EBL child than a dwelling with no poor LBP. No other frequency of lead-based painted components in any condition was found to be a significant predictor of dwellings with an EBL child.

Table 6.3.8: Paint Variables and the Significant Predictors of Children with Elevated Blood Lead Levels

Paint Variable	**Milwaukee**		**New York**	
	Predictor (P-value)	**Percent Increase in Odds of EBL with Failure**	**Predictor (P-value)**	**Percent Increase in Odds of EBL with Failure**
# Surfaces w/LBP	N	-	N*	-
# Surfaces w/LBP in Fair/Poor Condition	N	-	N	-
# Surfaces w/LBP in Poor Condition	≥1 (p=0.019)	126%	N*	-
# Surfaces in Fair/Poor Condition	≥ 20 (p=0.008)	122%	N	-
	≥ 30 (p=0.094)	52%		
# Surfaces in Poor Condition	≥ 2 (p=0.042)	95%	N	-

N = None of the predictors were found to be significant in expected direction
*Marginally significant predictor in “wrong” direction

In Milwaukee, the frequency of deteriorated paint, regardless of the level of lead in the paint, was found to be a significant predictor of blood lead outcome. When a lead-based paint hazard is defined as non-intact paint, the presence of 20 or more deteriorated surfaces was predictive of a “high-risk” dwelling ($p<0.01$) and the presence of 30 or more was marginally predictive (p=0.09). When a lead-based paint hazard was defined as "poor" paint condition, the presence of two or more these surfaces was predictive of a "high-risk" dwelling (p=0.04). No other frequency of paint deterioration on any paint was significant in New York City or Milwaukee.

As was found in the previous section, the number of surfaces with lead based paint in *poor* condition in New York City was marginally predictive (p=0.05) but in the "wrong" direction. Likewise the total number of leaded surfaces in any condition was also marginally significant (p=0.06) in New York, but in the wrong direction. These results were considered spurious and are therefore not reported in detail.

6.3.6.3 Discussion of Findings

The results of these analyses offer some support for the HUD Paint Film Classification System. A paint lead hazard defined as leaded paint in "poor" condition was a statistically significant predictor of homes of children with elevated blood lead levels in Milwaukee. This result corroborates similar results presented in the previous section for this site. Unfortunately, confounding factors in New York City precluded the use of this site to confirm the universality of this conclusion.

Likewise, while the results must be considered cautiously because they were only true in Milwaukee, a single leaded component in "poor" condition was in fact the most predictive paint lead variable when trying to identify dwellings likely to house EBL children. The hypothesis of this objective, that a dwelling will not be a "high-risk" dwelling unless multiple leaded components are deteriorated, is therefore not supported by this analysis. It must be emphasized, however, that if this finding could be replicated in another locale, it could only be applied to pre-1950 urban housing, because the post-1950 dwellings were not eligible for this analysis. For the population of pre-1950 dwellings, these results bolster the original guidance in the HUD *Guidelines* that a dwelling should be considered to have lead-based paint hazard when a single component has a lead paint level of 1 mg/cm^2 or greater and is "poor" as defined by the Paint Film Classification System.

This objective explored a second hypothesis that it is not necessary to test deteriorated paint for lead. It was theorized that if the occurrence of deterioration on any painted surface was predictive of "high-risk" dwellings, the cost of paint lead testing could be avoided. Unfortunately, the results were inconclusive. In Milwaukee, the presence of any deterioration was predictive, but not in New York City. Further examination of this issue in future studies would be beneficial.

6.4 Exploration of the Components of Variation in Dust Lead Sampling

6.4.1 Purpose

This objective investigated the variability of dust lead sampling results within rooms, within dwellings, across dwellings and across sites. Unlike the other objectives in this report, the applicability of variability measurements to the development of better risk assessment protocols may not be apparent to the general reader. However, results of variability analyses should contribute to the decisions about many of the basic dust sampling design issues, such as how many samples are needed, where samples should be collected and which dust lead standard is appropriate. An understanding of the variability of dust lead within a dwelling should help to identify a more appropriate dust lead sampling protocol.

6.4.2 Methodology

To establish estimates of variability of dust lead, three sets of analyses were conducted. Each of the sets of analyses is briefly described below. A more detailed description of the methodology is available in Appendix C.3.

Part I: Reliability Ratios

In a subset of dwellings in the study, side-by-side reliability samples were collected. The reliability sample collection procedures for the study are summarized in Section 2.2.4.

Side-by-Side dust samples in the home were used to estimate side-by-side variability for each sample type and site. Side-by-side variability measures the variability between dust lead loading of adjacent samples. Sources of side-by-side variability include: variability due to laboratory analysis, spatial variability and other effects such as different pressure to the wipe (technician effect). The results of the mixed modeling used for this analysis were used to calculate the reliability ratios and 95% lower bounds for them. Reliability ratios are interpreted similarly to correlation coefficients.

Part II: Estimates of Variation Attributable to Building, Room and Error

All dust samples in the home (except additional side-by-side samples) were used to estimate between building variability and combined estimates of room/error variability for each sample type and site. The room/error variability was not separable since only one sample was collected per room on a given sample type. The side-by-side models discussed in Part 1 were used to estimate between side-by-side variability and combined building/room variability. The building/room variability was not separable since only one room per building had side-by-side sampling on a given sample type. These two modeling approaches were combined to give estimates of the percent of variation attributable to: building, room and error.

The three estimates can be summarized as:

- % Building: Variability between buildings. Reflects factors affecting the ENTIRE building.
- % Room: Variability between rooms. Reflects factors affecting the room that are not building level effects (e.g., paint lead in a room).
- % Error: The rest of the variability not attributable to other sources.

Part III: Simulation of the Effect of Variability

Using the combined estimate of room/error variability from Part II, observations were randomly generated from a log-normal distribution with these estimates of variability and various specified "true" average dust lead levels. The details are presented in Appendix C.3. This analysis was based on the assumption that there is some "true" unobservable dust lead level in a dwelling on a given surface type. Each dust sampling location was assumed to be equally representative of the true "unobservable" dust lead level in the dwelling on that surface type.

For the sample mean and maximum based on 1, 2 and 4 samples per dwelling, the following errors are evaluated:

(i) Type I (False Positive) Error = the probability that the <u>sample statistic fails</u> the dust lead standard given that the "<u>true" lead level is below the standard</u>.
(ii) Type II (False Negative) Error = the probability that the <u>sample statistic passes</u> the dust lead standard given that "<u>true" lead level is above the standard</u>.

6.4.3 Findings and Discussion of Reliability Ratios Analyses

The reliability ratios were calculated on the basis of the side-by-side samples (Table 6.4.1). Window sill measurements had reliabilities ranging from 0.791 in Milwaukee to 0.887 in Baltimore County. Window trough reliabilities ranged from 0.854 in Baltimore County to 0.923 in New York City. For both these surfaces, the reliabilities were similar across sites and were generally above 0.8, which is considered to be indicative of good reliability. In Milwaukee, the reliabilities of floor measurements were similar to those reported for sills and troughs (0.885 for uncarpeted floors and 0.812 for carpeted floors). The only surfaces that had reliabilities below 0.8 were floors in Baltimore County and New York City.

The lower floor reliability ratios in Baltimore County and New York City may be attributable to two factors. First, the floor dust lead loadings were very low at these sites; geometric mean floor dust lead loadings were 3 and 2 $\mu g/ft^2$ for uncarpeted and carpeted floors, respectively, in Baltimore County, and 4 and 3 $\mu g/ft^2$ for uncarpeted and carpeted floors, respectively, in New York City (Table 6.4.2). At these lower levels (especially after log transformation), small absolute differences can make the variability of side-by-side samples appear higher than expected. Second, a fairly substantial percentage of floor dust results were below the laboratory limit of detection (2 µg/ft^2); 43 percent and 23 percent in Baltimore County and New York City, respectively. Because all non-detectable floor samples were set as 1.4 µg/ft^2 (2 µg/ft^2 divided by the square root of 2), the true variance of a side-by-side sample where one sample is detectable and the other is not is unknown. It is suspected that the floor reliability ratios were not as affected in Milwaukee because of that site's higher geometric mean floor dust lead loadings (11 and 6 µg/ft^2 for uncarpeted and carpeted floors, respectively) and limited number of floor samples that were below the laboratory detection limit (4 percent).

Table 6.4.1: Reliability Ratios (and 95% Lower Confidence Bounds) for Side-by-Side Samples

Sample Type	Surface Type	Reliability Ratio		
		Baltimore County	Milwaukee	New York City
Window Sill	-	0.887 (≥ 0.840)	0.791 (≥ 0.713)	0.814 (≥ 0.731)
Window Trough	-	0.854 (≥ 0.796)	0.878 (≥ 0.828)	0.923 (≥ 0.884)
Central Floor	Bare	0.689 (≥ 0.586)	0.885 (≥ 0.840)	0.761 (≥ 0.663)
	Carpet	0.427 (≥ 0.286)	0.812 (≥ 0.735)	0.738 (≥ 0.600)

Table 6.4.2: Geometric Mean Dust Lead Loadings (µg/ft^2) with 95% Confidence Intervals for Side-by-Side Samples

Sample Type	Surface Type	Baltimore County		Milwaukee		New York City	
		Number of Dwellings (Samples)	Geometric Mean	Number of Dwellings (Samples)	Geometric Mean	Number of Dwellings (Samples)	Geometric Mean
Window Sill	-	54 (162)	16 (14,19)	53 (159)	407 (329,503)	40 (120)	32 (24,43)
Window Trough	-	54 (162)	115 (89,150	53 (159)	9640 (6944,13383)	39 (117)	391 (274,557)
Central Floor	Bare	54 (162)	3 (2,3)	58 (174)	11 (9,12)	43 (129)	4 (3,5)
	Carpet	52 (156)	2 (2,3)	47 (141)	6 (5,7)	26 (78)	3 (2,3)

The results are supportive of the reliability of dust wipe sampling. The combined factors of spatial variability, technician effect and laboratory variability (as well as other sources of error) produced reliability ratios that were quite acceptable (i.e., generally above 80%). Furthermore, the differences between sites were relatively small, suggesting that the factors that are more site-specific, like technician effect, do not have a large impact on the results.

6.4.4 Findings and Discussion of Estimates of Variation Attributable to Building, Room and Error

The percentages of variability attributable to the three different sources of variability (building, room and error) are presented in Table 6.4.3. On window sills, the percentage of variability attributable to building was less than 50 percent at each of the sites. The variability attributable to the room was at least the same if not greater than the variability attributable to the building. The results suggest that there were sources of lead that created differences between the window sill dust lead loadings from room to room.

On window troughs, variability attributable to building was higher than the variability attributable to room in Baltimore County and Milwaukee, but not in New York City. While the percentage of variability that was attributable to building was greater than 50 percent only in Milwaukee (61%), the results suggest that the dust lead loadings from room to room are more stable in both Milwaukee and Baltimore County (at least in relation to building to building variability). In New York City, window troughs had more within dwelling variability.

On uncarpeted central floors, the percentage of variability attributable to building was greater than 50 percent at each of the sites. The variability attributable to the room was generally much lower than the variability attributable to the building. Unlike window sills, the uncarpeted central floor dust lead loadings were more stable from room to room. This finding appears to correspond with the window friction analysis findings that found that window friction had a significant effect on the window sill dust lead loading, but it was not a significant factor on floor dust lead loadings. Variation in window friction effects from room to room would be expected to create variability between window sill dust lead loadings, but not between floor dust lead loadings.

The distribution of variability across the three sources was very similar on uncarpeted floors and carpeted floors in Milwaukee. In Baltimore County and New York City, the results were not as clear because of the higher percentages of variability attributable to error (i.e., side-by-side variability). The error variability was also higher on uncarpeted floors at these two sites, but its effect on the building-room variability relationship was not as dramatic. As discussed in the previous section (Section 6.4.3), the low levels of floor dust lead loadings and the issues created by the relatively large number of samples below the detection limits pose problems for the interpretation of variability on carpeted (and to a lesser extent, uncarpeted) floors in both Baltimore County and New York City.

Table 6.4.3: Percent of Variance Attributable to Building, Room and Error

Sample Location	Surface Type	City	% Variability Attributable to		
			Building	Room	Error[1]
Window Sills	-	Baltimore Co	21	68	11
		Milwaukee	27	52	21
		New York	41	40	19
		All	59	33	8
Window Troughs	-	Baltimore Co	49	36	15
		Milwaukee	61	27	12
		New York	35	57	8
		All[2]	69	25	6
Central Floor	Bare	Baltimore Co	52	17	31
		Milwaukee	56	32	12
		New York	58	18	24
		All	71	16	13
Central Floor	Carpet	Baltimore Co	XX[3]	XX[3]	XX[3]
		Milwaukee	60	20	19
		New York	42	32	26
		All	73	4	23

[1] Error variability = Side-by-side variability
[2] Modeling assumptions were not met, therefore results are less reliable.
[3] Estimation impossible due to the high error (i.e., side-by-side) variance estimates (56%).

When results are presented separately by site, the variance attributable to building is expected to be smaller because the dust lead loadings tend to be fairly homogeneous within the site. When all sites are combined, the variance component for building increases, as dwellings in different sites will not be as homogeneous. This study assumes that the building variability within a site is a better point of comparison to assess variability from room to room.

A major assumption of a risk assessment is that through the use of dust sampling, a risk assessor can determine the “true” maximum or average dust lead exposure that a child will encounter. As variability between sampling locations increases, the number of samples needed to get a more precise estimate of the “true” maximum or average must also increase. Because windows sills have a relatively high level of variability within a dwelling, more samples (relative to other surface types) may be needed to produce a better estimate of the “true” dust levels. Floors have relatively low levels of variability and therefore may require fewer locations to be sampled in a dwelling to achieve the same level of accuracy as sills. (Interestingly, risk assessors in this study selected an average of five floors to one window sill sample, or just the opposite of what the variability estimates would suggest as appropriate.) Window trough variability estimates were less consistent across sites (e.g., lower than floors in Milwaukee, higher than window sills in New York City) so the results are harder to interpret.

6.4.5 Findings and Discussion of the Simulation of the Effect of Variability

Part III of these analyses generated Type I and Type II error estimates for each combination of site, surface type (uncarpeted and carpeted floors, window sills and window troughs), number of samples (1-5), and a prescribed set of dust lead standards. To simplify the presentation of these numerous results, a limited number of estimates are presented in Table 6.4.4. Estimates that represent significance levels 0.05, 0.10 and 0.20 are presented for each of the sites for floors and window sills. For comparative purposes, the effects of having only one or two samples collected in the dwelling are presented for floors in Milwaukee.

From the perspective of being most protective of a child's health, the upper uncertainty bounds in the table are of most interest. For example, the 95% upper uncertainty bound for a window sill dust sample of four rooms in New York City (see bottom row of table) was 940 $\mu g/ft^2$ when the window sill standard was 250 $\mu g/ft^2$. In practical terms, this means that if the "true" average lead level is 940 $\mu g/ft^2$, then there is a 5% chance that the sample mean will be below the standard of 250 $\mu g/ft^2$. More troubling, if the "true" lead level is 560 $\mu g/ft^2$, then there is a 20% chance that the sample mean will be below the standard of 250 $\mu g/ft^2$. These estimates are based on the good recovery rates achieved by the labs in this study. If the recovery rate is low, the variability effects can be compounded. For example, as reported above, when the "true" lead level is 940 $\mu g/ft^2$, there is a 5% chance that the sample mean will pass the standard. If the recovery rate were 79%, this would effectively reduce the "true" lead level to about 740 $\mu g/ft^2$ thus inflating the probability of incorrectly passing the standard to 10%.

Sample variability may be just as harmful to the interests of a property owner and the affordability of housing. Using the sampling characteristics in the example above with the 80% lower uncertainty bound, approximately 20% of the time a home with a "true" lead level of 175 $\mu g/ft^2$ would fail the standard of 250 $\mu g/ft^2$. In other words, 20% of the time a dwelling with a true level 30% lower than the standard will fail the standard due to variability.

While the table offers evidence that with additional samples in a dwelling errors are less likely, the example above demonstrates that even with four samples the rate of error can be high. If 40 $\mu g/ft^2$ was established as a "health-based" standard for floors, these results suggest that that it may be appropriate to set an "action-level" below that standard to take into account the variability and be truly health protective.

The high levels of variability for window sills (and window troughs) may help explain why these components were not predictive of blood lead outcomes in the analyses in Section 6.1. It also raises questions about whether any sampling plan including window sills can be predictive of risk.

Table 6.4.4: Estimates of Upper and Lower Uncertainty Levels by Study Site, Surface Type, and Number of Samples

City	Surface	Standard	# of Samples	Confidence Level					
				95%		90%		80%	
				Lower	Upper	Lower	Upper	Lower	Upper
Milwaukee	Floor[1]	40	1	17	145	21	120	28	85
			2	20	99	23	83	30	67
			4	24	73	27	65	32	53
Balt. Co	Floor[1]	40	4	28	58	30	55	34	50
Milwaukee	Floor[1]	40	4	24	73	27	65	32	58
New York	Floor[1]	40	4	25	68	28	60	32	53
Balt. Co	Sill	250	4	120	630	150	540	190	440
Milwaukee	Sill	250	4	80	1630	115	1250	180	840
New York	Sill	250	4	105	940	140	740	175	560

[1] Central dust sampling location, carpets and bare floors combined

6.5 Investigation of the Optimum Dust Lead Sampling Locations

6.5.1 Purpose

This objective considered the relationship between dust lead sampling locations and blood lead outcomes. As was discussed earlier in Section 6, the HUD *Guidelines* and the final Section 402/404 rules from EPA give the risk assessor much discretion when selecting dust sampling locations. Risk assessors are expected to sample locations where children under the age of six are likely to have contact, but the number and location (either within a dwelling or within a room) of the samples has never been prescribed.

Section 5.2.2.3 offered evidence that, at least in this study, the performance characteristics for risk assessments where the risk assessor selected the sampling locations were similar to those for risk assessments where the researchers selected the locations *a priori*. Of course, the dust sampling locations contributed just a small part to the overall outcomes of those risk assessment protocols. It is possible that if other elements of the risk assessment are improved, the selection of dust sampling locations could play a more significant role in the performance of a risk assessment. For practical reasons, risk assessors must be given latitude in their sampling location selection, but further guidance based on empirical evidence might aid the performance of the risk assessment.

This section of the report examines the most predictive sampling locations, but these results are not meant to stand-alone. The modeling conducted in this section will, by definition, limit the number of locations considered to be the most predictive. A sampling plan may be aided by including the samples identified here, but the plan should not necessarily be limited to them.

Two different levels of sampling locations were considered in this section: rooms and floor locations. Specification of a fixed set of rooms to sample is difficult since their power to predict risk is extremely dependent on the use patterns of the family. A living room for one family may be a special room that is rarely used, while for another family, it may be the center of family activities during all waking hours. The rooms identified here will represent the preferred sampling locations for the average family's use patterns.

Regulators have also considered the advantages and disadvantages of identifying set sampling locations on a floor. A central floor sample can be collected in any room and is a likely location for child's play, but it will be furthest away from friction/impact surfaces, it's loadings will most likely vary by daily movement, and it is most likely to be covered by an area rug. Entry floor samples can be collected in any room and should reflect dust lead deposited from friction on doors, but may also be affected by daily traffic. Window floor samples should reflect dust lead deposited from friction/impact on windows, but may not be accessible in all rooms. While decisions about floor sampling locations are often made for feasibility reasons, empirical evidence could direct risk assessors to a more predictive site. (In addition, if dust lead loadings actually vary by floor location, it would seem most appropriate to set the standard based on the location.)

This section neither addresses standards nor issues such as “is the maximum or average a better measure of risk,” but instead examines which sampling locations were most predictive of an EBL child. The results should contribute to the decision-making process when designing an improved and cost-effective dust sampling protocol.

6.5.2 Objective #7a: Determine dust sampling locations and sample types (using single surface samples) most predictive of children’s blood lead status?

6.5.2.1 Methodology

T-tests were used to compare the geometric mean dust lead loadings of dwellings housing a child with an elevated blood lead level with those of dwellings without an EBL child. The variables tested are displayed in Table 6.5.1. All interior dust sampling locations with adequate data were included; however, data for a second bedroom were not included since they were collected in a limited number of dwellings.

Table 6.5.1: List of Dust Sampling Locations Tested in Each Surface Type Model

All Floors*	Window Sills	Window Troughs	All Surfaces
Living Room Kitchen Bathroom Index Child’s Bedroom Playroom** Unit Entry	Living Room Kitchen Bathroom Index Child’s Bedroom Playroom**	Living Room Kitchen Bathroom Index Child’s Bedroom Playroom**	All surfaces used for the individual surface types

*A separate result was considered for entryway, central, window and perimeter samples for floors in all room locations except the unit entry.
**Playroom was identified in this study by the parent/guardian

Data in this analysis were restricted to the 147 eligible dwellings in Milwaukee and New York City. As discussed in detail in Section 5.2.1, including Baltimore County with its cross-sectional design and only a single child with an elevated blood lead level would likely lead to erroneous findings. Initially, the data from Milwaukee and New York City were to be analyzed together, but analyses conducted for Section 7 found that the relationship between environmental lead levels and blood lead outcomes was markedly different between the two sites. Combining the data from these two sites would not provide useful findings. Therefore, results for the two sites are presented separately.

The t-tests were used to identify dust sampling locations where the geometric mean dust lead loadings were at least marginally significantly different ($p<0.10$) in “high-risk” and “low-risk” homes. For interior surfaces that were at least marginally significant, an analysis was conducted examining the relationship between the dwelling unit mean dust lead loadings in “high risk” and “low-risk” dwellings using different combinations of rooms. For example, the relationships between the mean dust lead loadings on central floors in “high-risk” and “low-risk” dwellings were examined using results from the living room, kitchen, bedroom1, bath and unity entry in different combinations was examined. Results were considered statistically significant at a p-value<0.05.

6.5.2.2 Findings

Table 6.5.2a: Statistical Significance Between Dust Lead Loadings and Blood Lead Outcomes for Single-Surface Dust Lead Samples by Sample Type and Room Type (Milwaukee)

Location	# Samples- Not EBL	GM Dust Lead (95% CI) Not EBL ($\mu g/ft^2$)	# Samples- EBL	GM Dust Lead (95% CI) EBL ($\mu g/ft^2$)	P-value
Main Entry Floor(H/P)	31	74 (43,127)	35	113 (69,185)	0.263
Main Entry Floor(Int)	38	18 (12,27)	35	38 (28,52)	0.004**
LR Entry Floor	38	9 (7,11)	34	20 (15,26)	<0.001**
LR Perimeter Floor	37	6 (5,8)	33	14 (10,19)	0.001**
LR Central Floor	38	6 (5,9)	34	14 (10,21)	0.002**
LR Window Floor	37	10 (6,15)	33	21 (14,30)	0.012**
LR Sill	34	227 (131,391)	31	271 (141,523)	0.679
LR Trough	31	5,777 (2580,12934)	26	3,590 (1665,7737)	0.406
K Entry Floor	38	9 (7,12)	35	20 (15,26)	<0.001**
K Perimeter Floor	37	8 (6,11)	34	14 (11,19)	0.011**
K Central Floor	38	9 (7,12)	35	16 (13,21)	0.003**
K Window Floor	37	12 (8,16)	34	29 (21,40)	<0.001**
K Sill	36	183 (97,344)	34	203 (116,356)	0.807
K Trough	30	6,197 (2649,14496)	32	5,399 (2410,12095)	0.818
BA Entry Floor	37	11 (8,16)	35	17 (12,24)	0.105
BA Perimeter Floor	32	10 (8,14)	30	20 (13,31)	0.020**
BA Central Floor	37	8 (5,11)	35	12 (10,15)	0.049**
BA Window Floor	31	10 (6,17)	29	17 (11,28)	0.135
BA Sill	27	500 (236,1059)	28	793 (347,1766)	0.431
BA Trough	18	7,536 (2674,21239)	24	17,514 (8272,37082)	0.205
BR1 Entry Floor	38	9 (7,12)	34	19 (14,26)	0.001**
BR1 Perimeter Floor	38	9 (6,13)	33	16 (11,25)	0.044**
BR1 Central Floor	38	6 (5,8)	34	16 (11,24)	<0.001**
BR1 Window Floor	38	9 (6,13)	33	19 (13,28)	0.009**
BR1 Sill	35	216 (128,366)	32	358 (217,590)	0.178
BR1 Trough	34	4,451 (2049,9668)	29	5,914 (3014,11602)	0.590
Play Floor (at Entry)	38	8 (6,10)	35	17 (14,21)	<0.001**
Play Floor (at Per.)	37	7 (5,8)	34	12 (8,16)	0.007**
Play Floor (at Center)	38	6 (5,8)	35	12 (9,17)	0.002**
Play Floor (at Window)	36	9 (7,13)	34	16 (12,22)	0.031**
Play Sills	34	236 (137,405)	30	349 (178,685)	0.377
Play Troughs	31	4,910 (2147,11229)	25	3,347 (1518,7381)	0.515
Bldg Entry Floor	16	96 (50,185)	21	121 (75,195)	0.582
Hallway floor	16	38 (21,71)	21	53 (32,89)	0.428
Exterior	37	46 (28,76)	35	133 (92,193)	0.001**

LR=Living Room, K=Kitchen, BA=Bathroom, BR1=Index Child's Bedroom, Play=Playroom
*=p<0.10, **=p<0.05

Table 6.5.2b: Statistical Significance Between Dust Lead Loadings and Blood Lead Outcomes for Single-Surface Dust Lead Samples by Sample Type and Room Type (New York City)

Location	# Samples- Not EBL	GM Dust Lead (95% CI) Not EBL (µg/ft^2)	# Samples- EBL	GM Dust Lead (95% CI) EBL (µg/ft^2)	P-value
Main Entry Floor(H/P)	35	12 (9,17)	38	16 (11,22)	0.309
Main Entry Floor(Int)	35	6 (4,7)	38	6 (5,8)	0.700
LR Entry Floor	32	5 (4,6)	34	4 (3,5)	0.393
LR Perimeter Floor	27	3 (3,4)	27	3 (3,5)	0.777
LR Central Floor	32	3 (3,4)	34	4 (3,5)	0.674
LR Window Floor	27	6 (3,11)	27	5 (3,8)	0.667
LR Sill	25	46 (26,83)	20	13 (8,21)	0.029**w**
LR Trough	20	633 (302,1325)	20	249 (112,553)	0.101
K Entry Floor	35	4 (3,5)	37	3 (3,4)	0.104
K Perimeter Floor	34	4 (3,6)	36	3 (3,4)	0.097**w**
K Central Floor	34	4 (3,5)	38	3 (3,4)	0.357
K Window Floor	33	5 (4,6)	36	4 (3,5)	0.279
K Sill	27	29 (17,49)	32	27 (14,51)	0.860
K Trough	23	250 (128,487)	31	305 (143,652)	0.702
BA Entry Floor	35	4 (3,6)	39	4 (3,6)	0.889
BA Perimeter Floor	27	5 (3,7)	22	3 (2,4)	0.154
BA Central Floor	35	4 (3,5)	37	4 (3,5)	0.959
BA Window Floor	27	4 (3,6)	23	3 (2,5)	0.354
BA Sill	23	32 (17,60)	22	33 (17,63)	0.956
BA Trough	21	478 (285,803)	21	654 (302,1417)	0.514
BR1 Entry Floor	34	4 (3,5)	36	4 (3,4)	0.506
BR1 Perimeter Floor	29	4 (3,5)	30	3 (2,5)	0.904
BR1 Central Floor	33	3 (3,4)	36	3 (2,5)	0.961
BR1 Window Floor	29	4 (3,5)	30	3 (3,5)	0.899
BR1 Sill	22	33 (20,53)	25	34 (21,57)	0.881
BR1 Trough	20	338 (202,566)	23	174 (94,322)	0.113
Play Floor (at Entry)	33	5 (4,6)	38	4 (3,4)	0.120
Play Floor (at Per.)	27	4 (3,6)	33	3 (3,5)	0.342
Play Floor (at Center)	33	4 (3,5)	38	4 (3,5)	0.773
Play Floor (at Window)	27	4 (3,6)	33	4 (3,6)	0.847
Play Sills	23	50 (29,86)	28	19 (11,31)	0.014**w**
Play Troughs	19	429 (234,786)	27	223 (112,443)	0.168
Bldg Entry Floor	31	26 (16,42)	34	34 (23,51)	0.383
Hallway floor	29	12 (9,17)	29	14 (10,21)	0.557
Exterior	35	55 (38,81)	39	74(49,110)	0.305

LR=Living Room, K=Kitchen, BA=Bathroom, BR1=Index Child's Bedroom, Play=Playroom

*=$p<0.10$, **=$p<0.05$, **w**=significant, but in "wrong direction"

Thirty-five dust sampling locations were examined separately in Milwaukee and New York City (Tables 6.5.2a and 6.5.2b). They included 31 locations within the dwelling unit, 3 locations in the common area, and 1 location on the exterior. In Milwaukee, the exterior dust lead loadings were significantly different in "high risk" and "low-risk" dwellings, as were 19 of the 21 interior floor locations. Only the floor samples at the entry and under the window in the bathroom were not significantly different. In all cases, the dust lead loadings were higher in the homes with an EBL child. None of the window locations were significantly different, nor were the common area dust lead loadings different.

Of the 35 sample locations examined in New York City, 3 had at least marginally significantly different dust lead loadings for "high-risk" and "low-risk" homes (Table 6.5.2). However, all three of the locations were significantly different in the "wrong" direction; the geometric mean dust lead loadings on playroom sills, living room sills and kitchen perimeter floors were higher in homes without children with elevated blood lead levels than homes with an EBL child. No dust lead loading at any location was a significant predictor of homes where an EBL child resided in the enrolled dwellings in New York.

6.5.2.3 Impact of When Dust Sampling Locations are Combined

Throughout Sections 6 and 7 of this report, a sampling plan determined *a priori* by the researchers has been used to assess the predictive power of dust lead sampling. This plan, which is described in detail in Section 5.2.2.2, used central floor samples from the dwelling unit entry, the living room, kitchen and index child's bedroom for analysis. Since individual floor dust sample locations were significant predictors of blood lead status in multiple rooms and across all of the locations within a room, consideration was given to the impact of the study team's decisions.

When the household mean dust lead loadings as measured by the four locations discussed above were compared to samples taken from the room entries, perimeter and under the window, all locations were found to be significant predictors ($p<0.05$) of the enrolled child's blood lead status (Table 6.5.3). While there were some differences between the results by location that match expectations (e.g., locations near friction surfaces (entries and windows) had slightly higher dust lead loadings, these differences were not significantly different. When Unit Entries were removed from the floor average, the predictive power remained quite similar. Removing of the Unit Entry sample reduced the mean household dust lead loading between 16 and 38 percent.

Table 6.5.3: Statistical Significance Between Geometric Mean Dust Lead Loadings[1] and Blood Lead Outcomes for Floor Samples by Floor Sample Location (Milwaukee)

Floor Sample Location- w/ or W/o Unit Entry	GM Dust Lead-Not EBL (μg/ft^2) (95% CI) (n=38)	GM Dust Lead-EBL (μg/ft^2) (95% CI) (n=35)	GM Dust Lead -EBL/ GM Dust Lead-Not EBL	P-value
Central	12 (9,16)	25 (19,35)	2.1	0.001
Room Entry	14 (10,18)	28 (23,36)	2.0	<0.001
Perimeter	13 (9,18)	24 (17,32)	1.8	0.013
Window	16 (11,24)	32 (24,43)	2.0	0.007
Central w/o Unit Entry	8 (6,11)	18 (12,25)	2.3	0.001
Room Entry w/o Unit Entry	10 (8,13)	22 (17,28)	2.2	<0.001
Perimeter w/o Unit Entry	9 (7,12)	15 (11,22)	1.7	0.024
Window w/o Unit Entry	13 (9,19)	27 (19,37)	2.1	0.006

[1] Based on the arithmetic mean within the dwelling.

Additional analyses examining all combinations of floor sampling locations, including the possible addition of samples from the bath were considered. As an example, the findings for the central floor samples are presented on Table 6.5.4. As demonstrated on this table, the vast majority of sampling plans were highly significantly related to the blood lead status of the enrolled child. All results are presented in Appendix Table D14.

A few trends emerged from the full analysis:

- The addition of samples from the bathroom tended to reduce the predictive power of the household mean.
- The samples from the room entry floor and the central floor had more combinations with significance levels below 0.01 than samples from floors under windows and perimeter floors (87 and 81% versus 52 and 48%, respectively).
- Except for samples from the window floor, a sample based on the average of the living room and bedroom was as predictive or more predictive than the average of the four locations used in this report (Unit Entry, Living Room, Kitchen, and Child's Bedroom).

Table 6.5.4: Geometric Mean (GM) of the Dwelling Unit Mean *Central Floor* Dust Lead Loadings Using Different Combinations of Rooms in a Dwelling (Living Room, Kitchen, Bath, Child's Bedroom and/or Unit Entry) by Blood Lead Outcome (Milwaukee)

	Room Locations					**Units w/o EBL**		**Units w/EBL**		**Ratio** GM EBL/ GM Not EBL	**P-value**[1]
# Rms	Unit Entry	LR	K	BA	BR1	# Units	GM Dust Pb (95% CI)	# Units	GM Dust Pb (95% CI)		
1					Y	38	6 (5, 8)	34	16 (11,24)	2.7	<0.001
2			Y		Y	38	8 (7,10)	35	18 (12,25)	2.3	0.001
2		Y			Y	38	7 (5, 9)	35	16 (11,24)	2.3	0.001
3		Y	Y		Y	38	8 (6,11)	35	18 (12,25)	2.3	0.001
3	Y	Y			Y	38	12 (8,16)	35	27 (19,38)	2.3	0.001
4	Y	Y	Y		Y	38	12 (9,16)	35	25 (19,35)	2.1	0.001
1		Y				38	6 (5, 9)	34	14 (10,21)	2.3	0.002
2		Y	Y			38	8 (6,11)	35	16 (12,22)	2.0	0.002
2	Y				Y	38	13 (10,19)	35	30 (21,42)	2.3	0.002
3	Y		Y		Y	38	13 (9,18)	35	27 (20,37)	2.1	0.002
2	Y	Y				38	13 (9,19)	35	29 (21,40)	2.2	0.002
3	Y	Y	Y			38	13 (9,18)	35	26 (20,35)	2.0	0.002
1			Y			38	9 (7,12)	35	16 (13,21)	1.8	0.003
1	Y					38	18 (12,27)	35	38 (28,52)	2.1	0.004
2	Y		Y			38	15 (11,21)	35	29 (23,38)	1.9	0.004
4	Y	Y		Y	Y	38	12 (8,17)	35	24 (17,33)	2.0	0.004
5	Y	Y	Y	Y	Y	38	12 (9,16)	35	23 (17,31)	1.9	0.004
3			Y	Y	Y	38	9 (7,12)	35	16 (12,23)	1.8	0.005
2				Y	Y	38	8 (6,10)	35	15 (11,21)	1.9	0.006
4		Y	Y	Y	Y	38	9 (6,12)	35	17 (12,23)	1.9	0.006
3	Y			Y	Y	38	13 (9,18)	35	25 (18,35)	1.9	0.006
4	Y		Y	Y	Y	38	13 (9,18)	35	24 (18,33)	1.8	0.006
3		Y		Y	Y	38	8 (6,11)	35	16 (11,22)	2.0	0.007
3	Y	Y		Y		38	13 (9,18)	35	24 (18,32)	1.8	0.008
4	Y	Y	Y	Y		38	13 (9,18)	35	23 (18,30)	1.8	0.008
3	Y		Y	Y		38	14 (10,20)	35	25 (19,32)	1.8	0.011
2	Y			Y		38	15 (10,22)	35	27 (21,35)	1.8	0.012
3		Y	Y	Y		38	9 (6,12)	35	15 (12,20)	1.7	0.015
2			Y	Y		38	9 (7,12)	35	15 (12,19)	1.7	0.017
2		Y		Y		38	8 (5,11)	35	14 (10,18)	1.8	0.021
1				Y		37	8 (5,11)	35	12 (10,15)	1.5	0.049

[1] P-value of the test of equality between the geometric mean dust lead loadings for dwelling units with EBL children compared to dwelling units without EBL children.

6.5.2.4 Discussion of Findings

Consistent with results discussed in Section 6.1, no dust lead levels from individual sampling locations were predictive of blood lead outcomes of the children enrolled in New York City. Likewise, no dust lead levels from individual *window* sampling locations were predictive of blood lead status in Milwaukee. At the same time, almost all floor dust lead locations were individually significantly associated with the presence or absence of a child with an elevated blood lead level in Milwaukee.

In Milwaukee, the choice of floor sampling locations and combination of locations had little difference on the ability to assess risk. Almost all combinations of floor sampling locations were highly associated with the blood lead outcomes. The findings suggest that adding samples from a bathroom would not be beneficial. The findings also suggest that it is preferable to sample the living room and the child's bedroom from the room entry or central floor. However, the results from these floor sampling locations were not dramatically different than from other sampling locations, so these findings may not extend beyond the current study population.

The HUD Guidelines recommend that risk assessors interview families to identify a child's "play room." If the dwelling is vacant or the family is unavailable, the risk assessor is to take a sample from the living room. There was little difference between the predictive power of floor dust lead loadings from the play room versus the living room on blood lead status in Milwaukee. In fact, the p-values for the living room floor samples were equal or better to the play area floor samples suggesting that identifying the play room may not be necessary.

Although the choice of floor sampling locations do not appear to make a difference on the predictive power of the mean floor dust lead loadings, they may have an impact on the optimal standard. In Milwaukee, the Unit Entry floor samples were about twice as high as the interior floor samples. As discussed in Section 6.1, a floor sample including the Unit Entry in the average would perform differently against a given standard than a floor sample with the Unit Entry.

6.5.3 Objective #7a: Determine dust sampling locations and sample types (using composite samples) most predictive of children's blood lead status?

6.5.3.1 Methodology

The same methodology used to analyze the relationship between single-surface dust lead loadings and blood lead outcomes was used to analyze the relationship between composite dust lead loadings and blood lead outcomes. Details of the single-surface methodology are presented in section 6.5.2.1.

The following (field collected) composite samples were tested individually:

- Uncarpeted floors (Central)
- Uncarpeted floors (Window)
- Carpeted floors (Central)
- Carpeted floors (window)
- Window Sills
- Window Troughs

6.5.3.2 Findings

In Milwaukee, three of the six composite sampling locations were significant predictors of the child's blood lead status: uncarpeted and carpeted central floors and carpeted floors under windows (Table 6.5.5a). Neither of the composite window dust lead samples had significantly different loadings by blood lead outcome.

Table 6.5.5a: Statistical Significance Between Dust Lead Loadings and Blood Lead Outcomes for Composite Dust Lead Samples by Sample Type (Milwaukee) [1]

Location	# Samples (not EBL)	GM Dust Lead (95% CI) Not EBL ($\mu g/ft^2$)	# Samples (EBL)	GM Dust Lead (95% CI) EBL ($\mu g/ft^2$)	p-value
Central Floor-Uncarpeted	38	10 (7,13)	34	18 (13,24)	0.004**
Central Floor-Carpeted	33	6 (4,7)	29	11 (8,13)	0.010**
Window Floor-Uncarpeted	35	21 (13,34)	32	24 (18,34)	0.327
Window Floor-Carpeted	33	6 (5,9)	28	13 (9,19)	0.003**
Window Sill	36	356 (230,551)	34	467 (278,782)	0.556
Window Trough	36	8,970 (4801,16758)	33	14,166 (8350,24034)	0.244

*=p<0.10, **=p<0.05

[1] See Appendix Table D14 for more extensive details.

Table 6.5.5b: Statistical Significance Between Dust Lead Loadings and Blood Lead Outcomes for Composite Dust Lead Samples by Sample Type (New York City) [1]

Location	# Samples (not EBL)	GM Dust Lead (95% CI) Not EBL ($\mu g/ft^2$)	# Samples (EBL)	GM Dust Lead (95% CI) EBL ($\mu g/ft^2$)	p-value
Central Floor-Uncarpeted	33	4 (4,6)	36	5 (4,6)	0.683
Central Floor-Carpeted	12	3 (2,4)	15	2 (2,3)	0.985
Window Floor-Uncarpeted	32	6 (4,9)	30	6 (4,9)	0.684
Window Floor-Carpeted	8	3 (2,5)	10	2 (1,2)	0.060**w**
Window Sill	35	52 (31,88)	34	32 (22,48)	0.665
Window Trough	33	784 (448,1372)	34	417 (225,773)	0.071**w**

*=p<0.10, **=p<0.05, **w**=significant but wrong direction

[1] See Appendix Table D14 for more extensive details.

In New York City, two locations had at least marginally significantly different dust lead loadings for homes with and without children with elevated blood lead levels (Table 6.5.5b). Like the single-surface samples, composite dust lead loadings at these locations were significantly different in the “wrong” direction; the geometric mean dust lead loadings on uncarpeted floors under windows and on window troughs were higher in homes without children with elevated blood lead levels than homes with an EBL child. No dust lead loading at any location was a significant predictor of homes where an EBL child resided in the enrolled dwellings in New York.

6.5.3.3 Discussion of Findings

The findings presented in this analysis of composite samples concur with the findings of the single-surface sampling analyses. At both sites, there was no significant relationship between dust lead loadings from window sill or window trough samples and the presence of a child with an elevated blood lead. In Milwaukee, samples taken from floors generally are predictive of blood lead outcomes. The only sample that was not statistically significant was uncarpeted window floors. This result suggests that central floors may be a preferable sampling location to assess risk than areas under the window.

7.0 INVESTIGATION OF BETTER RISK ASSESSMENT PROTOCOLS

7.1 Introduction

In the original study analysis plan, it was expected that one or more of the current risk assessment protocols would be a statistically significant predictor of blood lead outcomes (EBL/non-EBL). The analysis plan called for using the predictive protocols as a starting point and then incorporating the findings from Section 6 to improve them. Surprisingly, neither the 1995 HUD Guidelines nor the current risk assessment protocols were statistically significant predictors of blood lead outcomes at the pre-1950 dwellings in this study (Section 5).

The analysis plan was revised to attempt to identify risk assessment protocols that were significantly predictive of the presence or absence of a child with an elevated blood lead level (≥ 10 µg/dL). Instead of starting with the current risk assessment protocols, an analysis was designed to examine each of the typical components of a risk assessment (paint, dust (including floors, sills and troughs), soil, as well as water) as possible components of an optimal protocol. The analysis followed the approach used in Section 5 (tests of independence and assessment of performance characteristics) to identify potential risk assessment protocol candidates.

7.2 Methods

To identify predictive risk assessment protocols, dwellings from Milwaukee and New York City were examined. Both sites had sufficient children with elevated blood lead levels to analyze the outcomes adequately. The environmental media that were considered as possible elements of the risk assessments were:

- Arithmetic Mean Dust Lead Loadings from:
  - Floors: (Central Floor-Living Room, Child’s Bedroom, Kitchen, plus Unit Entry)
  - Window Sills: (Living Room, Kitchen)
  - Window Troughs: (Living Room, Child’s Bedroom)
- Perimeter Soil Lead Concentration
- Play Area Soil Lead Concentration
- Water Lead Concentration (first draw)
- Frequency of Non-Intact Lead-Based Paint on Interior Surfaces
- Frequency of Non-Intact Lead-based Paint on Exterior Surfaces

The media selected were based, in part, on decisions made during the development of the current EPA/HUD risk assessment standards. For example, earlier sections of this report examined the relationship between household maximum and arithmetic mean dust lead loadings on a child’s blood lead status. This section, however, only explores the relationship between mean dust lead loading and the presence or absence of a child with an elevated blood lead level. The decision was based on the fact that EPA determined that the household mean dust lead loading would be used in risk assessments. Likewise, EPA and HUD determined that deteriorated lead-based paint would be defined as any non-intact paint, so that definition is retained here. The choice of the central area of the floor, as opposed to the entry, window or other perimeter location and the decision to exclude floor samples from unit entries were made at the discretion of the researchers.

The selection of a more predictive risk assessment protocol focused on whether certain media could be dropped from or added to the current protocols and whether alternate standards would be more predictive. The alternate standards that were considered are discussed in Section 7.2.2. Originally, the analysis plan called for considering all dwellings from Milwaukee and New York City (n=134) together. However, a preliminary examination of the environmental lead data identified different relationships between the environmental lead media and children's blood lead levels at the two sites.

7.2.1 Examination of Environmental Lead by Site

Descriptive statistics for the aforementioned environmental lead media as well as household maximums of floor, window sill and window trough dust lead loadings in Milwaukee and New York City are presented on Table 7.1.

A two-way ANOVA was run for each statistic to test two hypotheses:

1. Controlling for site, the geometric mean level of the statistic is the same for homes where a child with an EBL resides as homes without a child with an EBL.
2. Controlling for the blood lead outcome (EBL/not EBL), the geometric mean level of the statistic is the same for Milwaukee and New York City.

Among the results of these analyses:

- Housing units in New York and Milwaukee had significantly different environmental lead levels. Only water lead levels were similar at the two sites.
- Only mean floor dust lead loadings and perimeter soil lead concentrations were related to the blood lead status. Both the maximum and arithmetic mean floor dust lead loading for a dwelling were significantly different in the two sets of dwellings in Milwaukee. However, mean floor dust lead loadings were not significantly different between the two sets of dwellings in New York. (Water lead concentration was the only other environmental lead media that had a p-value below 0.3.)
- Window sill and trough lead loadings were not related to blood lead status. Furthermore, the results went in the "wrong" direction in New York where dust lead loadings were higher in dwellings without a child with an elevated blood lead level. This was also true for number of surfaces with non-intact interior lead-based paint and play area soil lead in New York City.

It was also observed that soil lead was often unavailable at these sites. Play area soil was infrequently available from both sites: 39 percent and 6 percent of dwellings in Milwaukee and New York City, respectively, had play area soil collected. Only 25 percent of the dwellings in New York City had perimeter soil available.

Table 7.1: Descriptive Statistics of Environmental Lead Media by Blood Lead Outcome (EBL/Non-EBL) and Site

Statistic	Site	N	N w/ EBL	Lead Levels (GeoMean)[1]		Test EBL=Non EBL	Test ML=NY
				EBL Homes	Non-EBL Homes		
Floor Dust Lead (max) (µg/ft^2)	ML	64	31	45	23	No p=0.03	No P<0.01
	NY	69	36	8	7		
Floor Dust Lead (mean) (µg/ft^2)	ML	64	31	24	12	No p=0.02	No P<0.01
	NY	69	36	4	4		
Sill Dust Lead (max) (µg/ft^2)	ML	62	33	459	355	Yes p=0.84	No P<0.01
	NY	63	32	36	52		
Sill Dust Lead (mean) (µg/ft^2)	ML	62	30	299	247	Yes p=0.69	No P<0.01
	NY	63	32	28	43		
Trough Dust Lead (max) (µg/ft^2)	ML	59	27	6,749	5,171	Yes p=0.67	No p<0.01
	NY	55	28	239	422		
Trough Dust Lead (mean) (µg/ft^2)	ML	59	31	9,601	6,895	Yes p=0.79	No p<0.01
	NY	55	28	282	483		
Perimeter Soil Lead (ppm)	ML	56	30	2,918	1,298	No p<0.01	No p<0.01
	NY	17	32	965	457		
Play Area Soil Lead (ppm)	ML	25	14	287	261	Yes p=0.86	No p=0.05
	NY	4	3	773	948		
Water Lead (first draw) (ppb)	NY	64	31	3	3	Yes p=0.15	Yes p=0.13
	ML	69	36	4	3		
Number of LBP Surfaces-Non-Intact (Exterior)	ML	64	31	6	8	Yes p=0.33	No p<0.01
	NY	69	36	1	1		
Number of LBP Surfaces-Non-Intact (Interior)	ML	64	31	18	14	Yes p=1.00	No p<0.01
	NY	69	36	4	7		

[1] For Number of LBP Surfaces-Non-Intact, the arithmetic mean values are presented and tested instead of the geometric mean values.

7.2.2 Final Analytical Design

The wide variation in lead levels in Milwaukee and New York City raised questions about how enrolled children were being exposed to lead at each site. The lack of relationships between lead levels in the homes in New York City and the children's blood lead status suggested that the home environment in New York was not a primary source of poisoning. As further support, if most children in New York City had elevated blood lead levels because of their housing environment, then the number of children with blood lead levels would have been much higher in Milwaukee. It was assumed that the home environment was a primary source of lead exposure in Milwaukee, given the observed relationships between certain environmental lead media and blood lead levels in that site. Therefore, it was hypothesized that any paint risk assessment protocol would fail to be predictive of blood lead status for the children enrolled in New York City, while alternative risk assessment protocols would be predictive of blood lead outcomes in Milwaukee.

The analytical design was adjusted so that the predictive power of a series of alternative risk assessment protocols was first examined using the only Milwaukee data. If predictive protocols for Milwaukee were identified, these protocols would be applied to dwellings in New York City. An "accurate" protocol for both New York City and Milwaukee would have a relatively high specificity and positive predictive value at both sites, but would have a relatively high sensitivity and negative predictive value only in Milwaukee. If protocols with these characteristics could be identified, then it would suggest that the causal hypotheses were correct.

Examining all permutations of the environmental lead media and standards listed in Table 7.2 created a list of possible predictive protocols. A total of 92,190 protocols were generated. For reporting purposes, each protocol was assigned a unique number.

Table 7.2: Environmental Lead Media and Standards Examined

Media	Standards Examined					
Floor Dust Lead (mean) (µg/ft^2)	None	10	15	25	40	100
Sill Dust Lead (mean) (µg/ft^2)	None	125	250	500		
Trough Dust Lead (mean) (µg/ft^2)	None	800	5,000	10,000		
Perimeter Soil Lead (ppm)	None	400	1,200	2,000	5,000	
Play Area Soil Lead (ppm)	None	400	(1,200 was tested but no sample was above this level)			
Water Lead (first draw) (ppb)	None	5	10	15		
Number of LBP Surfaces-Non-Intact (Exterior)	None	1	5	10		
Number of LBP Surfaces-Non-Intact (Interior)	None	1	5	10		

Although the ANOVA analyses suggest that only floor dust lead, perimeter soil lead and water lead had even a slight relationship with blood lead outcome, all environmental media were included as candidate protocols under the assumption that the non-significant media might have had effects at more extreme values or in combination. When an environmental lead sample was not available from a dwelling (such as soil lead in New York City), the dwelling was assumed to have passed the risk assessment for that media at all standards.

The performance characteristics for protocols that were at least marginally significantly associated ($p<0.10$) with blood lead status in Milwaukee were examined. The methodology used was the same as was described in Section 5.1 with the exception that the fixed dust sampling locations defined in Section 7.2 were used instead of the sampling locations identified by the risk assessor. For presentation in this report, the list of protocols was then culled to those protocols using the following procedure:

a. For each level of sensitivity, the sampling combination with the highest specificity was selected. If there was >1 protocol with the same sensitivity and specificity, all were selected.
b. For each level of sensitivity and specificity, the sampling combination with the highest level of association with EBL status (i.e., lowest p-value) was selected. If there was >1 protocol with the same sensitivity, specificity, and p-value, all were selected.
c. Finally if *both* the sensitivity *and* the specificity are lower than those for another sampling combination, then the combination was dropped.
d. The list of selected protocols was ranked by sensitivity.

7.3 Findings

7.3.1 Results of Analysis of Milwaukee Data

The analysis of the 64 Milwaukee dwellings identified 20 risk assessment protocols that optimized the sensitivity and specificity (Tables 7.3 & 7.4). These protocols are defined as the *optimal protocols*. As a general rule of thumb, protocols with a high *sensitivity* would be most protective of children's health, while a high *specificity* would be protective of either the affordable housing stock (if property owners pay for treatment) or the government's limited housing budget (if the government pays for treatment).

Table 7.3: Standards for Optimal Protocols in Milwaukee[1]

RANK	Protocol #	Floor Dust Pb (µg/ft^2)	Window Sill Dust Pb (µg/ft^2)	Window Trough Dust Pb (µg/ft^2)	Perimeter Soil Pb (ppm)	Play Area Soil Pb (ppm)	Interior Pb Paint (# not intact)	Exterior Pb Paint (# not intact)	Water Pb (ppb)
1	15941	10	-	-	2,000	400	-	-	-
2	15937	10	-	-	2,000	-	-	-	-
3	15365	10	-	-	-	400	-	-	-
4	15361	10	-	-	-	-	-	-	-
5	31493	15	-	-	5,000	400	-	-	-
6a	30725	15	-	-	-	400	-	-	-
6b	31489	15	-	-	5,000	-	-	-	-
6c	77383	100	-	-	2,000	400	-	-	10
7a	583	-	-	-	2,000	400	-	-	10
7b	77381	100	-	-	2,000	400	-	-	-
8a	581	-	-	-	2,000	400	-	-	-
8b	77379	100	-	-	2,000	-	-	-	10
9a	579	-	-	-	2,000	-	-	-	10
9b	77377	100	-	-	2,000	-	-	-	-
10	577	-	-	-	2,000	-	-	-	-
11	46851	25	-	-	5,000	-	-	-	10
12	46849	25	-	-	5,000	-	-	-	-
13	46083	25	-	-	-	-	-	-	10
14	46081	25	-	-	-	-	-	-	-
15	62209	40	-	-	5,000	-	-	-	-

[1] See Appendix Table D5 for more extensive details.

Table 7.4: Performance Characteristics for Optimal Protocols in Milwaukee (sorted by Sensitivity) [1]

	Protocol	% Fail	P-Value	Sensitivity	Specificity	PPV	NPV
01	#15941	81	0.000	100 (89,100)	36 (20,55)	60 (45,73)	100 (74,100)
02	#15937	78	0.001	97 (83,100)	39 (23,58)	60 (45,74)	93 (66,100)
03	#15365	73	0.000	94 (79,99)	45 (28,64)	62 (46,75)	88 (64,99)
04	#15361	70	0.001	90 (74,98)	48 (31,66)	62 (47,76)	84 (60,97)
05	#31489	59	0.006	77 (59,90)	58 (39,75)	63 (46,78)	73 (52,88)
06a	#30725	59	0.006	77 (59,90)	58 (39,75)	63 (46,78)	73 (52,88)
06b	#31493	64	0.002	84 (66,95)	55 (36,72)	63 (47,78)	78 (56,93)
06c	#77383	59	0.006	77 (59,90)	58 (39,75)	63 (46,78)	73 (52,88)
07a	#00583	56	0.006	74 (55,88)	61 (42,77)	64 (46,79)	71 (51,87)
07b	#77381	56	0.006	74 (55,88)	61 (42,77)	64 (46,79)	71 (51,87)
08a	#00581	53	0.007	71 (52,86)	64 (45,80)	65 (46,80)	70 (51,85)
08b	#77379	53	0.007	71 (52,86)	64 (45,80)	65 (46,80)	70 (51,85)
09a	#00579	50	0.012	68 (49,83)	67 (48,82)	66 (47,81)	69 (50,84)
09b	#77377	50	0.012	68 (49,83)	67 (48,82)	66 (47,81)	69 (50,84)
10	#00577	47	0.012	65 (45,81)	70 (51,84)	67 (47,83)	68 (49,83)
11	#46851	44	0.011	61 (42,78)	73 (54,87)	68 (48,84)	67 (49,81)
12	#46849	41	0.010	58 (39,75)	76 (58,89)	69 (48,86)	66 (49,80)
13	#46083	34	0.035	48 (30,67)	79 (61,91)	68 (45,86)	62 (46,76)
14	#46081	31	0.030	45 (27,64)	82 (65,93)	70 (46,88)	61 (45,76)
15	#62209	27	0.048	39 (22,58)	85 (68,95)	71 (44,90)	60 (44,74)

[1] See Appendix Tables D6 and D7 for more extensive details.

7.3.2 Discussion of Findings from Milwaukee

Certain factors emerged from the results in Milwaukee:

- Floor dust lead loadings and perimeter soil lead concentrations were the two exposure sources most likely to be included in the most predictive protocols. These findings reinforce the earlier findings that these media were most predictive of the presence or absence of a child with an elevated blood lead level.

- The optimal protocols included the complete range of mean floor dust lead loading standards tested. They also included the higher levels of perimeter soil lead concentrations tested (2,000 and 5,000 ppm).[6]
- Some of the optimal protocols included play area soil lead (400 ppm) and water lead (10 ppb). While the play area level matches the current standard, the water lead level is 5 ppb lower than the current action level.
- Window sill and window trough dust lead and frequency of interior and exterior non-intact lead-based paint were not elements of the 20 optimal protocols. These results match the earlier findings that these media were not predictive of homes in this study with or without a child with an elevated blood lead level.

7.3.3 Results of Analysis of New York City Data

Dwellings in New York City were assessed using the Milwaukee protocols. As hypothesized, none of these housing risk assessments were predictive of blood lead status (EBL/not EBL). Of the 20 protocols, only five had a p-value below 0.50 (Table 7.5).

As discussed earlier, if the hypothesis is correct that children enrolled in this study in New York City tended to be exposed to lead from sources other than the housing environment, then the lack of predictive power is not surprising. The goal of an optimal assessment tool for these homes would be to maximize the specificity and positive predictive power. Seven protocols had a specificity above 90 percent and a positive predictive value at or above 60 percent. Because no perimeter soil lead concentration in New York City exceeded 5,000 ppm and no mean floor dust lead loading exceeded 100 ug/ft2, essentially four protocols met these criteria:

1) (#46081) floor dust lead: 25 µg/ft^2
2) (#46083) floor dust lead: 25 µg/ft^{2}and water lead:10 ppb
3) (#62209) floor dust lead: 40 µg/ft^2
4) (#581) perimeter soil lead: 2,000 ppm and play area soil lead: 400 ppm

7.3.3 Discussion of Findings from New York City

Although these results appear to support the protocols that have mid-range floor dust lead loadings (25 or 40 ug/ft2) or mid-range soil lead concentrations, the number of dwellings with lead levels in these ranges was quite limited in New York City. Of the 69 dwellings studied in New York, just a handful failed these protocols (Table 7.6). In fact, the protocol with the highest failure rate for Milwaukee (#15941 – 81%) only had a failure rate of 19 percent in New York City. With just 13 dwellings failing this protocol in New York, the confidence interval around the positive predictive value is fairly large (25-81%). The New York City data appear to lack the discriminatory power to identify optimal standards for floor dust lead, soil lead or even water lead.

---

[6] The current risk assessment protocols call for a Rest of Yard soil lead sample instead of a Perimeter soil lead sample. If the Rest of Yard sample is collected equally from the perimeter of the building and an area similar to the play area, then a perimeter standard of 2,000 ppm averaged with a play area standard of 400 ppm would be equivalent to the current Rest of Yard standard of 1,200 ppm.

Table 7.5: Performance Characteristics for Optimal Protocols in New York [1]

Protocol	% Fail	P-Value	Sensitivity	Specificity	PPV	NPV
01 #15941	19	1.000	19 (8,36)	82 (65,93)	54 (25,81)	48 (35,62)
02 #15937	14	1.000	14 (5,29)	85 (68,95)	50 (19,81)	47 (34,61)
03 #15365	16	1.000	17 (6,33)	85 (68,95)	55 (23,83)	48 (35,62)
04 #15361	12	1.000	11 (3,26)	88 (72,97)	50 (16,84)	48 (35,61)
05 #31489	9	1.000	8 (2,22)	91 (76,98)	50 (12,88)	48 (35,61)
06a #30725	13	1.000	14 (5,29)	88 (72,97)	56 (21,86)	48 (35,62)
06b #31493	13	1.000	14 (5,29)	88 (72,97)	56 (21,86)	48 (35,62)
06c #77383	16	1.000	17 (6,33)	85 (68,95)	55 (23,83)	48 (35,62)
07a #00583	16	1.000	17 (6,33)	85 (68,95)	55 (23,83)	48 (35,62)
07b #77381	7	1.000	8 (2,22)	94 (80,99)	60 (15,95)	48 (36,61)
08a #00581	7	1.000	8 (2,22)	94 (80,99)	60 (15,95)	48 (36,61)
08b #77379	13	1.000	14 (5,29)	88 (72,97)	56 (21,86)	48 (35,62)
09a #00579	13	1.000	14 (5,29)	88 (72,97)	56 (21,86)	48 (35,62)
09b #77377	3	1.000	3 (0,15)	97 (84,100)	50 (1,99)	48 (35,60)
10 #00577	3	1.000	3 (0,15)	97 (84,100)	50 (1,99)	48 (35,60)
11 #46851	13	0.481	17 (6,33)	91 (76,98)	67 (30,93)	50 (37,63)
12 #46849	3	0.494	6 (1,19)	100 (89,100)	100 (16,100)	49 (37,62)
13 #46083	13	0.481	17 (6,33)	91 (76,98)	67 (30,93)	50 (37,63)
14 #46081	3	0.494	6 (1,19)	100 (89,100)	100 (16,100)	49 (37,62)
15 #62209	3	0.494	6 (1,19)	100 (89,100)	100 (16,100)	49 (37,62)

[1] See Appendix Table D7 for more extensive details.

Table 7.6: Frequency of Dwellings Failing Possible Standards in New York City

Media	Level	# of Dwellings[1] At Level	# of Dwellings w/EBL Child
Mean Floor dust lead loading	$\geq 40\ \mu g/ft^2$	2	2 (100%)
Mean Floor dust lead loading	25-39 “	0	-
Mean Floor dust lead loading	15-24 “	4	1 (25%)
Mean Floor dust lead loading	10-14 “	2	1 (50%)
Perimeter soil lead	≥ 2,000 ppm	2	1 (50%)
Play Area soil lead	≥ 400 “	3	2 (67%)
Water lead (first draw)	≥ 10 ppb	7	4 (57%)

[1] Out of 69 dwellings

7.4 Discussion

By analyzing alternative protocols, a series of protocols were identified that were more predictive of blood lead outcomes than the current protocols. The findings suggested floor dust lead and soil lead added value to the risk assessments and window sill and trough dust lead and observations of non-intact lead-based paint did not add predictive value. In some cases, water lead tests appeared to add predictive value as well.

This study demonstrates broad differences in environmental lead levels in Milwaukee and New York City even though these dwellings were all built prior to 1950. This study assumed that the impact of the environmental lead level of any given media on a child's lead exposure would not vary from community to community. However, the different outcomes for Milwaukee and New York City suggested that some children in New York City were getting poisoned from a source other than their housing environment. An examination of other sources of lead exposure is presented in Appendix A.

The different lead levels observed at the different sites underscores the need for additional studies of the relationship between environmental lead levels and childhood blood lead levels in multiple communities. Unlike this study, the Rochester Lead-in-Dust Study found that window sill dust lead was related to children's blood lead. Since the Rochester study was an important study in the development of the current risk assessment protocols and standards, it is logical that tests of the window sill are part of the current protocols. However, the findings of this study raise questions regarding whether window sill dust lead[7] is more generally related to blood lead levels.

Although this study included 134 children living in pre-1950 housing, the disparity in environmental lead conditions did not support a combined analysis of homes in Milwaukee and New York City. As a result, the power of the study was not as robust as originally expected. While the findings of this study strongly support sampling floor dust lead loadings and soil lead concentrations, the exact standards that should be set for these media are less clear.

The findings supported further examination of the current protocols. As presented in Table 7.7, modifications to the current protocols, including changing the definition of paint deterioration, or dropping paint and sill tests improved the predictive power of these risk assessments. By dropping paint observations and window sill dust testing, the modified current protocols were marginally predictive in New York City (p=0.08) and almost marginally predictive in Milwaukee (p=0.10). Finally, by raising the Perimeter soil lead standard to 2,000 ppm so that it better matched the current Rest of Yard soil lead standard (see footnote 6) the predictive power and the specificity improved in Milwaukee, while the specificity was left unchanged in New York City.

---

[7] An early decision for these analyses was to designate two rooms for window sill and window trough dust lead sampling. After completing the analyses, questions were raised if sill dust lead or trough dust lead would have been a significant predictor of EBL status had a sample from a third room been included. Additional investigation found that adding the third room did not improve the predictive power of these two media.

Table 7.7: Performance Characteristics of Current Protocols and Modified Current Protocols[2]

Protocol	Site	% Fail.	P-Value	Sensitivity	Specificity	PPV	NPV
Current Protocol[1]	Mil.	100	-	100 (89,100)	XX	48 (36,61)	XX
	NYC	99	0.478	100 (90,100)	3 (0,16)	53 (40,65)	100 (3,100)
Current – Except Paint Deteriorated If Poor	Mil.	94	0.114	100 (89,100)	12 (3,28)	52 (38,65)	100 (40,100)
	NYC	72	0.600	69 (52,84)	24 (11,42)	50 (36,64)	42 (20,67)
Current - Except No Paint Testing	Mil.	80	0.062	90 (74,98)	30 (16,49)	55 (40,69)	77 (46,95)
	NYC	26	0.422	31 (16,48)	79 (61,91)	61 (36,83)	51 (37,65)
Same as Above Plus No Sill Dust Testing	Mil.	70	0.104	81 (63,93)	39 (23,58)	56 (40,70)	68 (43,87)
	NYC	14	0.087	22 (10,39)	94 (80,99)	80 (44,97)	53 (39,66)
Same as Above, Plus. Perimeter=2000 ppm	Mil.	58	0.013	74 (55,88)	58 (39,75)	62 (45,78)	70 (50,86)
	NYC	10	0.431	14 (5,29)	94 (80,99)	71 (29,96)	50 (37,63)

[1] Dwelling fails Current Protocols if results at or above following action levels: Floor Dust Lead: 40 µg/ft^2; Window Sill Dust Lead: 250 µg/ft^2; Play Area Soil Lead: 400 ppm; Perimeter Soil Lead: 1,200 ppm; Any non-intact lead-based paint on interior or exterior

[2] See Appendix Table D4 for more extensive details.

Table 7.7 suggests that the current protocols could be made a fairly predictive tool by dropping two media without changing standards. Further study may conclude that changes to the current standards could further improve the risk assessment protocols. Yet, this study supports the premise that environmental lead results can be used to identify homes where children are likely to have elevated blood lead levels.

8.0 REFERENCES

[1] U.S. Department of Housing and Urban Development: Guidelines for the Evaluation and Control of Lead-Based Paint in Housing. HUD, Washington, DC (1995).

[2] Housing Environmental Services, Housing Authority Risk Retention Group: Lead-based Paint Risk Assessment Protocols. HES, Cambridge, Massachusetts (1991).

[3] U.S. Department of Housing and Urban Development: Requirements for notification, Evaluation and Reduction of Lead-Based Paint Hazards in Federally Owned Housing and Housing Receiving Federal Assistance. 24CFR § 35 *Fed Reg* 64(178):50140-50231 (1999).

[4] U.S. Environmental Protection Agency. Lead: Identification of Dangerous Levels of Lead. 40CFR § 745 *Fed Reg* 66(4):1205-1240 (2001).

[5] U.S. Environmental Protection Agency. Lead-Based Paint Risk Assessment Model Curriculum, U.S. EPA, Washington, DC (1995).

[6] JetForm (1997-1998). The data entry for this paper was implemented using JetForm Filler and JetForm Design, Version 2.23.2. Copyright © 1997-1998. JetForm Corporation. JetForm, JetForm Design and JetForm Filler are registered trademarks of JetForm Corporation. Ottawa, Ontario, Canada.

[7] SAS Institute (1989-1996). The data analysis for this paper was generated using SAS/STAT software, Version 6.12 of the SAS System for Windows. Copyright © 1989-1996. SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks of SAS Institute Inc., Cary, NC, USA.

[8] U.S. Centers for Disease Control and Prevention: Update: Blood-Lead Levels – United States, 1991-1994 *MMWR* 46(7):141-6 (1997).

[9] Last, JM (Ed.). Dictionary of Epidemiology. Oxford University Press, New York, New York (1983).

[10] Lanphear BP; Burgoon DA; Rust SW; et al.: Environmental Exposures to Lead and Urban Children's Blood Lead Levels. *Environ. Res.* 76:120-130 (1998).

[11] Lanphear, BP; Matte, TD; Rogers, J; et al.: The Contribution of Lead-Contaminated House Dust and Residential Soil to Children's Blood Lead Levels: A Pooled Analysis of 12 Epidemiologic Studies. *Environ Res* 79:51-68 (1996).