Appendix 12: Statistical Rationale for Sample Sizes and Percentages Used in Guidance for Inspecting in Multi-Family Housing

This appendix presents the statistical rationale and calculations used to develop sample sizes (number of units to be tested) in multi-family housing. (See Note 1, below.) The samples sizes apply both to inspections for lead-based paint and to post-abatement dust clearance testing in multi-family housing. The appendix also presents the detection capability of the sampling scheme, that is, the probability that the scheme will successfully detect various levels of contamination in the housing development tested.

12A.1 Sample Size Calculations

To determine the applicable sample size using the methods of this appendix, the housing units must first be properly grouped. For lead-based paint inspections, similar units and buildings should be grouped based on common construction, floor plans, and painting history. This type of grouping will make it less likely that lead-based paint will be missed in the testing. Likewise, for dust clearance testing, units and buildings that have similar construction and were cleaned in the same manner should be grouped for sampling purposes.

Because the sampling scheme applies to both testing for lead-based paint inspections and dust clearance testing, the term "the HUD standard" will be used to mean either 1.0 mg/cm^2 for lead-based paint inspections or the applicable clearance standard for dust testing. The term "component" means a floor, windowsill or window well for dust clearance testing, and means any painted building component for lead-based paint inspections.

The basic specification for the sampling scheme is that it achieves 95% confidence that at least 95% of the units meet the HUD standard. This means that, if all units sampled meet the HUD standard for all components tested, there is 95% confidence that fewer than 5% of the units in the development have one or more components in violation of the HUD standard, assuming no with-in unit sampling error and no measurement error. An alternate interpretation is that up to a 5% chance of missing some lead in up to 5% of the units is allowed. In a large development, 5% of the units might be a large absolute number, however, so the total number of leaded units that might escape detection has been limited to 50. This leads to the following quantitative prescription for the sampling plan:

TEST THE SMALLEST NUMBER OF UNITS WITH THE PROPERTY THAT, IF ALL TESTED UNITS ARE AT OR BELOW THE HUD STANDARD FOR ALL COMPONENTS, THERE IS 95% CONFIDENCE THAT THE NUMBER OF UNITS WITH AT LEAST ONE COMPONENT AT OR ABOVE THE HUD STANDARD IS LESS THAN 50 UNITS OR 5%, WHICHEVER IS SMALLER.

As an example, 56 units should be tested in a 600-unit development. Sample sizes where taken from Table 7.3 of Chapter 7 (or Table IV in this appendix). If no lead (above the HUD standard) is found in any of the 56 tested units, the owner of the development can be 95% confident that less than 30 units (the lesser of 50 and 5% of 600) have lead above the HUD standard. As a second example, 232 units should be tested in a 4000 unit development. If all are below the dust clearance HUD standard for all tested floors, windowsills and window troughs, there is 95% confidence that less than 50 of the 4000 units (the lesser of 50 and 5% of 4000) have any lead dust levels at or above the applicable HUD standard. Note that developments with 20 or fewer units, all units should be tested and the classification rules for single-family housing apply.

The statistical calculations required to determine the number of units to be tested, based on the criterion above, are fairly straightforward. For the sake of brevity, call a unit with one or more components with lead-based paint (or dust lead, as the case may be) at or above the HUD standard a "leaded unit." Make the following definitions:

N = Total number of units in the development;

k = Maximum allowable number of leaded units;

n = Smallest number of units which must be tested to provide 95% confidence that the total number of leaded units is "k" or less, based on finding no leaded units in the sample tested.

For example, if 95% confidence is required that less than 5% of 300 units have lead, then k = 14. If 95% confidence is required that fewer than 50 units have lead out of 4000 units is required, then k = 49.

In the usual statistical convention, "n" is defined as the smallest integer for which the probability of obtaining no positive results in a simple random sample of size "n" from a population of size "N", of which k+1 are positive, is less than 0.05. (See Note 2, below.) When k+1 of "N" total are positive, the probability of observing no positive results in a simple random sample of size "n" is given by the hypergeometric formula[1]:

$[(N-k-1)\ldots(N-k-n)]/[(N)(N-1)\ldots(N-n+1)]$.

The required value of "n" is obtained by successively evaluating this expression for n = 1, 2, 3,..., until its value first drops below 0.05. The calculations were performed in SAS[2], using the hypergeometric distribution function[1]. Table I shows the exact values of "k" and "n" for selected values of "N".

In developing the sample sizes for Table 7.3 of Chapter 7, two refinements to the calculations were made. First, because of the discrete nature of the problem, it is possible for the sample size to decrease when the total number of units increases. To see how this happens, suppose that a building has 40 units. Since 5% of 40 is two, the maximum number of leaded units allowed is 1. However, if the building has 41 units, 5% of 41 is 2.05, so the maximum number of leaded units is 2. Since it is obviously easier to detect 2 units out of 41 than 1 out of 40, the minimum sample size for a building with 41 units is smaller than the minimum sample size for a building with 40 units. Specifically, the exact sample size for 40 units is 31, while the exact sample size for 41 units is 26. The same problem occurs every time the number of units is a multiple of 20. Since it is extremely counter-intuitive for the sample size to decrease when the number of units increases, the additional requirement that the sample size never decrease was imposed. The result of this requirement, which can be observed clearly in Table 7.3, is that the sample size remains constant for some time beginning at each multiple of 20.

The second refinement to the calculation was to calculate a percentage of units to be sampled when the total number of units is very large. When the total number of units is 1,000 or greater, the maximum acceptable number of leaded units is 49. Suppose that a proportion "P" of the N units is to be tested when N is large. Then, when the number of leaded units is 50, the minimum unacceptable number, the probability that zero leaded units will be found in the sample can be approximated by (1-P)50 = 0.05 if P=0.058. (The ratio of n to N in Table I is approximately 0.058 for N greater than 1000). Thus, the limiting percentage for the sample size is 5.8%. In Table 7.3, the sample size is taken as 5.8% of the number of units, rounded to the nearest whole number, when N is 1,040 or larger.

12A.2 Detection Capability of the Sampling Scheme

By the detection capability of the sampling scheme is meant the probability that the sample contains at least one leaded unit when leaded units are present. Thus, the detection capability is the probability that a problem (lead-based paint or dust above the applicable HUD standard) will be detected in the development, in the sense of showing up in at least one of the units in the sample.

The detection capability of the sampling scheme depends on the total number of leaded units in the development as a whole. Clearly, the more leaded units there are, the better the chance that they will appear in the sample. When the number of leaded units is k+1 (in Table I), the detection capability is, by definition, (slightly) greater than 95%. In general, when the number of leaded units is "L", the detection capability is calculated from the formula

$$1 - [(N-L)(N-L-1)...(N-L-n+1)]/[(N)(N-1)...(N-n+1)]$$

where "N" and "n" are, respectively, the total number of units in the development, and the sample size. Table II shows the number of leaded units that must be present in the development as a whole for the detection capability to be 50%, 75%, 90%, 95%, 97.5%, or 99%.

Table I. Calculation of Number of Units to Be Tested In Multi-family Developments.

Calculation of the Number of Units to be Tested		
N[a]	k[b]	n[c]
20	0	20
40	1	31
60	2	38
80	3	42
100	4	45
200	9	51
300	14	54
400	19	55
600	29	56
1,000	49	57
1,500	49	86
2,000	49	115
2,500	49	144
3,000	49	174
3,500	49	203
4,000	49	232
4,500	49	261
5,000	49	290

[a]**N** = Number of Units in the Development;
[b]**k** = Maximum Allowable Number of Leaded Units;
[c]**n** = Number of Units to be Tested

Tables II and III give probabilities of finding at least one leaded unit in the tested sample. This does not mean that all, or even most, of the leaded units will be sampled. To achieve this would require virtually 100% sampling. The expected percentage of the leaded units that will be sampled is equivalent to the sampling percentage, i.e., the sample size as a percentage of the number of units in the development. For example, in a 100-unit development, 45 units are sampled (highlighted in yellow in Table I). Thus, 45% of the leaded units would also be expected to be sampled, on average. In a 1,000-unit development, an average of 5.7% of the leaded units would be sampled.

Table II. Calculation of Number of Leaded Units Which Must Be Present in the Development as a Whole for Various Levels of Detection Capability.

		Detection Capability					
Na	nb	50%	75%	90%	95%	97.5%	99%
20	20	1	1	1	1	1	1
40	31	1	1	2	2	3	3
60	38	1	2	3	3	4	5
80	42	1	2	3	4	5	6
100	45	2	3	4	5	6	8
200	51	3	5	8	10	12	15
300	54	4	7	12	15	18	23
400	55	5	10	16	20	25	30
600	56	7	14	23	30	37	45
1,000	57	12	24	39	50	61	75
1,500	86	12	24	39	50	61	76
2,000	115	12	24	39	50	61	76
2,500	144	12	24	39	50	62	76
3,000	174	12	24	39	50	62	76
3,500	203	12	24	39	50	62	76
4,000	232	12	24	39	50	62	76
4,500	261	12	24	39	50	62	76
5,000	290	12	24	39	50	62	76

[a]**N** = Number of Units in the Development;
[b]**n** = Number of Units Tested

As an example, the detection capability of the scheme in a 600-unit development is 99% when the development contains 45 leaded units (highlighted in yellow in Table II). This means that the sample of 56 units in a 600-unit development is 99% certain to include at least one of the 45 leaded units. Notice that the numbers are almost exactly the same for all developments with 1,000 units or more. This reflects the design decision that the number of leaded units which may be missed completely (with 5% probability) must be less than 50. Of course, the fixed numbers in the table reflect a decreasing percentage of the total number of units in the development. Table III shows the percentage of leaded units that must be present to achieve the various detection capabilities.

Table III. Calculation of Percentage of Leaded Units Which Must Be Present to Achieve the Various Detection Capabilities.

		Detection Capability					
N[a]	n[b]	50%	75%	90%	95%	97.5%	99%
20	20	5.0%	5.0%	5.0%	5.0%	5.0%	5.0%
40	31	2.5%	2.5%	5.0%	5.0%	7.5%	7.5%
60	38	1.7%	3.3%	5.0%	5.0%	6.7%	8.3%
80	42	1.3%	2.5%	3.8%	5.0%	6.3%	7.5%
100	45	2.0%	3.0%	4.0%	5.0%	6.0%	8.0%
200	51	1.5%	2.5%	4.0%	5.0%	6.0%	7.5%
300	54	1.3%	2.3%	4.0%	5.0%	6.0%	7.7%
400	55	1.3%	2.5%	4.0%	5.0%	6.3%	7.5%
600	56	1.2%	2.3%	3.8%	5.0%	6.2%	7.5%
1,000	57	1.2%	2.4%	3.9%	5.0%	6.1%	7.5%
1,500	86	0.8%	1.6%	2.6%	3.3%	4.1%	5.1%
2,000	115	0.6%	1.2%	2.0%	2.5%	3.1%	3.8%
2,500	144	0.5%	1.0%	1.6%	2.0%	2.5%	3.0%
3,000	174	0.4%	0.8%	1.3%	1.7%	2.1%	2.5%
3,500	203	0.3%	0.7%	1.1%	1.4%	1.8%	2.2%
4,000	232	0.3%	0.6%	1.0%	1.3%	1.6%	1.9%
4,500	261	0.3%	0.5%	0.9%	1.1%	1.4%	1.7%
5,000	290	0.2%	0.5%	0.8%	1.0%	1.2%	1.5%

a[N] = Number of Units in the Development;
b[n] = Number of Units Tested

For buildings or groups of similar buildings with 1,040 units or more, test 5.8 percent of the number of units, rounded to the nearest unit. EXAMPLE: If there are 2,170 units, 5.8 percent times the number of units is 125.86 units, so 126 units should be tested.

For example, in a 1,000-unit development, the detection capability of the scheme is 75% when 2.4% of the units are leaded (highlighted in yellow in Table III). That is, the sample of 57 units tested is 75% sure to contain at least one of the 2.4% of units that are leaded. Put another way, although 97.6% of the units have no lead above the HUD standard, a random sample of only 5.7% of the units has a 75% chance of finding one of the leaded units. The percentages in Table III are fixed (except for round off) for developments with 200–1,000 units, and then decline for larger developments. Again, this is the result of the design decision to fix the percentage of leaded units that may be missed (with 5% probability), for developments with 1,000 units or less. For larger developments, the number of such units is fixed, but the percentage is declining.

12A.3 Sample Size and Decision Percentages in the Multi-family Decision Flowchart

To obtain 99% confidence on conclusions made about a component type using the multi-family decision flowchart in Chapter 7, XRF readings must be taken on at least 40 components of the given type. A sample size of 40 was chosen as a minimum sample size that could be achieved in almost all cases given that at least 20 units would be tested in a multi-family housing development.

For simplicity, a single percentage was desired for declaring a component type either positive or negative in multi-family housing. The decision rule in the flowchart to declare a component type positive is based on the percentage of XRF readings classified as positive relative to the HUD standard and the decision rule to declare a component type negative is based on the percentage of XRF readings less than the HUD 1.0 mg/cm^2 standard, assuming a 5% false positive rate and a sample size of at least 40. Parameters provided in the *XRF Performance Characteristics Sheet* for each specific XRF instrument were developed so that the false positive rate would be 5%. Thus, for sample sizes of 40 or greater and when operating an XRF instrument as specified in the *XRF Performance Characteristics Sheet*, 99% confidence may be obtained for the following:

- ✦ At least one component of a given type has lead in paint equal or greater than the HUD standard if 15% of the components are classified as positive relative to the HUD standard.
- ✦ None of the components of a given type have lead in paint greater than the HUD standard if 100% of the XRF readings taken on the components of a given type are less than 1.0 mg/cm^2.

The statistical rationale for the percentages used in the decision rules of the flowchart is given below.

Positive Percentage in Multi-family Decision Flowchart

The Multi-family Decision Flowchart (Figure 7.1 of Chapter 7) gives the following rule: based on XRF readings, if 15% or more components of a given type are classified as positive relative to the HUD standard, then the inspector concludes that lead is present at 1.0 mg/cm^2 or greater on at least one of the components of the type tested. Assuming a true false positive rate of 5%, the 99th percentiles of the observed number and percentage of false positive classifications for several sample sizes are shown below in Table IV.

With a sample size of at least 40 for a component type and if the components all have true lead levels less than the HUD standard (1.0 mg/cm^2), there is only a 1% probability of observing 15% or more positive results. In other words, if 15% or more results are actually observed on a component type, one can be 99% confident that lead is present on at least one of the components of a given type. Since 15% is the percentage that corresponds with a sample size of 40, 15% was adopted as the cutoff percentage for declaring a component type positive relative to the HUD standard in Chapter 7.

Table IV. Number and Percentage of False Positive Classifications for Several Sample Sizes Assuming True False Positive Rate of 5%

Sample Size	Number of False Positive Results	Percentage of False Positive Results
20	4	20
40	6	15
60	7	12
80	9	11
100	11	11

Negative Percentage in Multi-family Decision Flowchart

The flowchart specifies that if 100% of the XRF readings taken on components of a given type are less than 1.0 mg/cm^2, the conclusion is that no lead is present at or above the 1.0 mg/cm^2 HUD standard on the component type.

Given that the sample size must be at least 40 (as described above), suppose that exactly 1 of the 40 components tested has true lead level of 1.0 mg/cm^2 or greater. Then, the probability of obtaining an XRF reading less than 1.0 mg/cm^2 on all (100%) of the components of the given type is:

Pr(All XRF readings < 1.0 mg/cm^2) =

Pr(1 true lead level ≥1.0 mg/cm^2 has XRF reading < 1.0 mg/cm^2)

times Pr(39 true lead levels < 1.0 mg/cm^2 have XRF readings < 1.0 mg/cm^2) = $p_1 \times p_2^{39}$,

where

p_1 = probability a true lead ≥ 1.0 mg/cm^2 has XRF reading < 1.0 mg/cm^2

p_2 = probability a true lead < 1.0 mg/cm^2 has XRF reading < 1.0 mg/cm^2

The maximum value of this expression using results from XRF instruments examined by EPA in a large field study [3a, 3b] was 0.017. Thus, if one or more of the 40 components is truly positive (lead level 1.0 mg/cm^2 or greater) relative to the HUD standard, there is less than a 2% chance of obtaining XRF readings less than 1.0 mg/cm^2 on all (100%) components of the component types. This means that, whenever all XRF readings on a component of a given type are less than 1.0 mg/cm^2, there is at least 98% confidence that none of the 40 components have true lead above 1.0 mg/cm^2.

With the application of the flowchart and with a sample size of 40, there is a very high probability (at least 98 percent) that a tested component type will be correctly classified. Combined with the 95 percent probability that at least one leaded component will be selected for inspection by the sampling scheme described above when 5 percent or more of the components have lead-based paint at or above 1.0 mg/cm^2, the procedure provides an overall confidence level of between 93 percent and 95 percent.

12A.4 Sample Size as a Function of Multifamily Development Size for 1960-1977 Developments

For 1960-1977 building developments, a similar procedure is followed except the quantitative prescription would be to: Test the smallest number of units with the property that, if all tested units are at or below the HUD standard for all components, there is 95% confidence that the number of units with at least one component at or above the HUD standard is less that 100 units or 10%, whichever is smaller. The SAS[2] program used to perform this calculation for pre-1960 and 1960-1977 multi-family buildings with 10 to 1040 units is below.

For 1960-1977 building developments, when the total number of dwelling units is 1,000 or greater, the maximum acceptable number of leaded units is 99. Suppose that a proportion "P" of the units is to be tested when "N" is large. Then, when the number of leaded units is 100, the minimum unacceptable number, the probability that zero leaded units will be found in the sample can be approximated by (1-P)100=0.05 if P=0.029. (The ratio of n to N in Table 7.3 is approximately 2.9%. In Table 7.3, the sample size is taken as 2.9% of the number of units, rounded to the nearest whole number, when N is 1,040 or larger.)

```
****************************************************************;
**Sample Size as a Function of Multi-family Development Size ***;
**[SAS Program]************************************************;
****************************************************************;
title 'Basis For Table 7.3 of the HUD Guidelines';
****************************************************************;
****************************************************************;
**Output Variables: ***;
** capn = # units in building ***;
** k1= Pre 1960 Max Allowable # leaded units ***;
** k2= 1960-1977 Max # Allowable leaded units ; ***;
** nc05=Pre 1960 # units to test in a building of size "capn"***;
** nc10=1960-77 # units to test in a building of size "capn" ***;
****************************************************************;
**Other (working) Variables: ***;
** n = # units tested in a building of size "capn" ***;
```

```
** r1=probability of obtaining no positive results in a ***;
** sample of size "n" from a population of size "capn" ***;
** of which "k1+1" are positive ***;
** r2=probability of obtaining no positive results in a ***;
** sample of size "n" from a population of size "capn" ***;
** of which "k2+1" are positive ***;
*****************************************************************;
data set1;
**loop over all possible size buildings*************************;
 do capn=10 to 1040;
 label capn='# units in building';
**determine the maximum allowable # of leaded units*************;
 k1=ceil(capn*0.05)-1;
 if k1>49 then k1=49;
 label k1='Pre 1960 Max # leaded units';
 k2=ceil(capn*0.10)-1;
 if k2>99 then k2=99;
 label k2='1960-1977 Max # leaded units';
**determine the # of units to sample for pre 1960 units*********;
 n=1; r1=1;
****loop through until r1 falls below 0.05 *********************;
 * do while (round(r1,.000000001)>=.0500 and n<=capn);
 do while (r1>=.0500 and n<=capn);
 if (capn-(k1+1)>=n) then r1=probhypr(capn,(k1+1),n,0);
 nc05=n;
 n=n+1;
 end;
**determine the # of units to sample for 1960-1977 units********;
 n=1; r2=1;
****loop through until r2 falls below 0.05 *********************;
* do while (round(r2,.0001)>=.0500 and n<=capn);
 do while (r2>=.0500 and n<=capn);
 if (capn-(k2+1)>=n) then r2=probhypr(capn,(k2+1),n,0);
 nc10=n;
 n=n+1;
 end;
 output;
 end;
label nc05='Pre 1960 # units to test';
```

```
label nc10='1960-1977 # units to test';
run;
title2 'Estimates Based on the Hypergeometric Distribution';
proc print label noobs;
 var capn k1 nc05 k2 nc10;
run;
****Add the non-decreasing sample-size requirement**************;
data set2;
 set set1;
 if _N_=1 then do;
 q05=nc05;
 q10=nc10;
 end;
 else do;
 q05=max(nc05,q05);
 q10=max(nc10,q10);
 end;
 retain q05 q10;
run;
proc sort data=set2;
 by q05 q10;
run;
proc means noprint;
 var capn;
 by q05 q10;
 output out=out2 min=min5 max=max5;
run;
data out2;
 set out2;
 if min5^=max5 then range5=compress(min5||"-"||max5); else range5=put(min5,5.);
run;
title2 'Estimates for Table 7.3 (adjusted to be non-decreasing)';
proc print label noobs;
 var range5 q05 q10;
label q05='Pre 1960 # units to test';
label q10='1960-1977 # units to test';
label range5='# units in building';
run;
```

NOTES:

1. The primary contributor to the programming, data analyses and preliminary drafting of this Appendix 12 was Sherry L. Dixon, Ph.D., of the National Center for Healthy Housing, whose work on this statistical rationale is appreciated.

2. k+1 is used to determine the probability for at most k positive values. This assures that the occurrence of k positive values will have probability less than 0.05.

References

[1] W. Feller (1968). *An Introduction to Probability Theory and its Applications, Volume I*, Third Edition, Wiley, New York. ISBN: 978-0-471-25708-0. ***http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471257087.html***

[2] SAS Institute Inc., SAS Campus Drive, Building T, Cary, NC 27513-8617. ***http://www.sasinstitute.org/index.shtml***

[3a] D. C. Cox, F. G. Dewalt, M. M. Haugen, R. A. Koyak, R. L. Schmehl. A field test of lead-based paint testing technologies: Summary report (May 1995). U.S. Environmental Protection Agency contract EPA-68-CO-0137. ***http://www.epa.gov/lead/pubs/summary.txt***

[3b] D. C. Cox, F. G. Dewalt, M. M. Haugen, R. A. Koyak, R. L. Schmehl. A field test of lead-based paint testing technologies: Technical report (May 1995). U.S. Environmental Protection Agency contract EPA-68-CO-0137. Available at cost from the National Technical Information Service, ***http://www.ntis.gov/***, document number PB96125026.

[The websites were accessed 7/28/2012; they may be moved or deleted later.]