DRAFT TECHNICAL REPORT

on

MODELS FOR NATIONAL LEAD LABORATORY ACCREDITATION PROGRAM (NLLAP) EXPANSION

May 27, 1999

Prepared by

Steven M. Bortnick, Abi Katz-Stein,
Peter A. Chace, and Ann M. Herberholt

BATTELLE
505 King Avenue
Columbus, Ohio 43201-2693

for

John Scalera, Work Assignment Manager

Office of Pollution Prevention and Toxics
U.S. Environmental Protection Agency
Washington, D.C. 20460

EPA Contract No. 68-W-99-033
Work Assignment 1-4

DISCLAIMER

The material in this document has not been subject to Agency technical and policy review. Views expressed by the authors are their own and do not necessarily reflect those of the U.S. Environmental Protection Agency. Mention of trade names, products, or services does not convey, and should not be interpreted as conveying, official EPA approval, endorsement, or recommendation. Do not quote or cite this document.

This reports is copied on recycled paper.

TABLE OF CONTENTS

Page

EXECUTIVE SUMMARY v

1.0 INTRODUCTION 1
- 1.1 BACKGROUND 1
- 1.2 SCOPE OF REPORT 1
- 1.3 OBJECTIVE OF REPORT 2
- 1.4 ORGANIZATION OF REPORT 2

2.0 DEFINITIVE LABORATORY/ANALYSIS 4
- 2.1 DEFINE AND SPECIFY REQUIREMENTS FOR DEFINITIVE CLASSIFICATION 4
- 2.2 ANALYSIS OF ELPAT DATA FOR NLLAP RECOGNIZED LABORATORIES 5
- 2.3 PRESENT AND DEFINE A DECISION TREE FOR 95 PERCENT CONFIDENCE 5

3.0 SEMI-QUANTITATIVE LABORATORY/ANALYSIS 7
- 3.1 CHARACTERIZE XRF PRECISION 7
  - 3.1.1 XRF Performance from Characteristic Sheets 7
  - 3.1.2 EPA Field Study on XRF Measurement Precision 10
  - 3.1.3 Field Investigation of On-Site Techniques (Portable XRF) 11
  - 3.1.4 Using XRF Technology for Soil Analysis 11
- 3.2 CHARACTERIZING ULTRASONIC EXTRACTION/ANODIC STRIPPING VOLTAMMETRY (UE/ASV) PRECISION 12
  - 3.2.1 EPA Evaluation of the PaceScan 2000 12
  - 3.2.2 Interlaboratory Evaluation of UE/ASV Lead Measurements on Paint, Dust, and Soil 13
  - 3.2.3 Field Investigation of On-Site Techniques (UE/ASV) 14
  - 3.2.4 Laboratory Evaluation of the PaceScan 2000 14
  - 3.2.5 Other Studies Considered 15
- 3.3 PRESENT AND DEFINE A DECISION TREE FOR 95 PERCENT CONFIDENCE 16

4.0 QUALITATIVE LABORATORY/ANALYSIS 18
- 4.1 USING QUALITATIVE ANALYSES AS NEGATIVE SCREENS 18
- 4.2 USING QUALITATIVE ANALYSES AS POSITIVE SCREENS 20
- 4.3 FIELD TEST RESULTS FROM EVALUATING CHEMICAL TEST KITS 22
  - 4.3.1 EPA Field Study on Chemical Test Kits 22
  - 4.3.2 Field Investigation of On-Site Techniques (Chemical Test Kits) 23
- 4.4 PRESENT AND DEFINE A DECISION TREE FOR 95 PERCENT CONFIDENCE, USING BOTH A NEGATIVE SCREEN AND A POSITIVE SCREEN 23

5.0 ADDITIONAL ISSUES AND CONCERNS 26
- 5.1 AN ALTERNATIVE DECISION TREE FORMAT 26
- 5.2 LABORATORY VERSUS TESTING FIRM DISTINCTION 27
- 5.3 DIFFICULTIES IN ASSESSING PERFORMANCE 28
- 5.4 REPEATED SAMPLING TO IMPROVE PRECISION 28

TABLE OF CONTENTS (Continued)

Page

5.5 ALTERNATIVES TO THE GRAY-ZONES PROVIDED IN THIS REPORT 29
5.5.1 An Alternative Approach that Avoids Gray-Zone Calculations..... 29
5.5.2 An Alternative Gray-Zone Calculation 29
5.6 CONCLUSION 30

6.0 REFERENCES 31

APPENDIX A: GLOSSARY A-1

APPENDIX B: DETAILS FOR DATA ANALYSES CONDUCTED..... B-1

List of Tables

Table 2.1. Gray-zone comparison between definitive laboratory requirements and NLLAP recognized laboratories..... 5
Table 3.1. Gray-zones (mg/cm^2) for various XRF instruments when measuring lead levels on painted surfaces, based on PCS data..... 9
Table 3.2. Gray-zones defined as 1.0 ± two times the precision in mg/cm^2, where 0 is the lowest limit. Results are based on the above described field test data..... 10
Table 3.3. PaceScan 2000 results based on data from (EPA 600/EPA-95/093), April 1996..... 13
Table 3.4. Results based on data from interlaboratory evaluation of UE/ASV..... 14
Table 3.5 Results based on data from analysis of TCLP Extracts (PaceScan 2000)..... 15
Table B.1 Regression parameters for an NLLAP accredited laboratory's precision as a function of the true lead level..... B-1
Table B.2 Descriptive statistics for paint, dust, and soil sample means for full and reduced data sets..... B-2
Table B.3. Estimated regression parameters for the response component and the SD component..... B-6

List of Figures

Figure 2.1. Decision tree for making statements with 95 percent confidence using a Definitive Laboratory..... 6
Figure 3.1. Decision tree for making statements with 95 percent confidence using a Semi-Quantitative Laboratory..... 17
Figure 4.1. Hypothetical operating characteristic (OC) curve of a chemical test kit analyzing lead in paint (demonstrating qualitative analysis performance considered appropriate as a negative screen)..... 19
Figure 4.2. Hypothetical operating characteristic (OC) curve of a chemical test kit analyzing lead in paint (demonstrating qualitative analysis performance considered appropriate as a positive screen)..... 21
Figure 4.3. Decision tree for making statements with 95 percent confidence, using both a negative and positive screen..... 25
Figure B.1. Relationship between overall standard errors and means for NLLAP recognized laboratory analyses of paint, dust, and soil..... B-3

EXECUTIVE SUMMARY

This report presents options and issues associated with the expansion and redesign of the current National Lead Laboratory Accreditation Program (NLLAP) to cover laboratories and lead testing firms generating data for the evaluation of potential lead hazards from paint chips, dust, and soils. A field-decision performance based model for providing 95 percent confidence in decision making is developed for laboratories and testing firms engaged in at least one of three types of analysis: definitive, semi-quantitative, and qualitative. Definitive analyses are reflective of the NLLAP's current laboratory quality system requirements (LQSR). Semi-quantitative analyses are defined as quantitative analyses that produce data which do not meet the performance requirements for a definitive analysis but are still of sufficient quality to support a decision at the lead level of concern. Qualitative analyses do not provide quantitative information but are still capable of determining the presence or absence of lead.

Every analyzing instrument has associated measurement error. When a recorded lead level is near the action level, the uncertainty associated with measurement creates a "gray-zone" in decision making. This gray-zone is a band around the action level where the true concentration of lead cannot be judged, with a required amount of certainty, to be above or below the action level, due to imprecision of the measuring instrument. A definitive laboratory is quantitatively defined as a laboratory having a confirmed gray-zone that is no larger than plus or minus 20 percent of the action level of concern. Analyses of ELPAT data for NLLAP recognized laboratories from rounds 14-21 show that on average, these laboratories are able to perform within the prescribed standard of the action levels plus or minus 20 percent for paint chips, dust, and soil. In other words, the performance of NLLAP recognized laboratories meets the requirement for definitive laboratories. The field-decision performance based model of this report recommends that results lying within the gray-zone of a definitive laboratory be classified as positive for lead above the action level. This conclusion is conservative and protects those at risk for exposure to lead.

Analyses of available data for field-portable X-ray fluorescence (XRF) instruments and ultrasonic extraction/anodic stripping voltammetry (UE/ASV) instruments used to detect lead were performed. Results show that laboratories using XRF instruments tend to perform at the semi-quantitative level near the action level of concern. Laboratories using UE/ASV instruments sometimes gave results that could be classified as definitive, and sometimes as semi-quantitative near the action level of concern. The field-decision performance based model of this report recommends that semi-quantitative laboratory results that fall within the laboratory's gray-zone be sent to a definitive laboratory for a more accurate confirmatory analysis. Any results outside the gray-zone indicating the presence or absence of lead are accepted as accurate with 95 percent confidence.

Laboratories or testing firms using chemical test kits that indicate only a presence or absence of lead are classified as qualitative laboratories. It is noted that current test kits available are useful as either a negative screen (i.e., whether the amount of lead is below the action level) or a positive screen (i.e., whether the amount of lead is above the action level), but not both. The field-decision performance based model of this report recommends classifying negative results

from negative screens as below the action level, and positive results from positive screens as above the action level. The samples associated with other qualitative results must be sent to a definitive laboratory for a more accurate confirmatory analysis. A decision rule is given for cases when both negative and positive screens are available. If both screens agree, the result may be accepted as accurate with 95 percent confidence. It is recommended that if the screens reach opposite conclusions, the sample should be sent to a definitive laboratory for further analysis, since a decision cannot be reached with 95 percent confidence in this case.

In summary, a field-decision performance based model is recommended for providing 95 percent confidence in decision making. The form of the model will depend on the type of analysis being performed: definitive, semi-quantitative, or qualitative.

1.0 INTRODUCTION

1.1 BACKGROUND

In the FY92 appropriations bill, Congress identified the Environmental Protection Agency (EPA) as the federal agency responsible for establishing an accreditation program for laboratories participating in the analysis of lead in paint chips, soils, and dust wipes, as part of a national home lead-based paint abatement and control program. The Office of Pollution Prevention and Toxics (OPPT) has established the National Lead Laboratory Accreditation Program (NLLAP) to help assure parties utilizing the services of laboratories recognized by NLLAP that the laboratories are capable of adequately performing lead analysis.

NLLAP recognition of laboratories analyzing lead in paint chips, soils, and dust wipes has two requirements: (1) Successful participation in proficiency testing using real world matrices, and (2) laboratory accreditation including on-site assessment of laboratory operations. The Environmental Lead Proficiency Analytical Testing (ELPAT) program is designed to administer this proficiency testing and assessment program. Design of the NLLAP is based on the recommendations of a Federal Interagency Taskforce on Lead-Based Paint, a group of 17 federal agencies involved with lead issues, that recognition should be based upon both proficiency testing and laboratory accreditation.

Currently the NLLAP applies to laboratories performing analysis on collected samples using quantitative methods. Laboratories or testing firms which perform analysis directly on the area in question (on site, *in-situ*) or use methodologies which produce data of less accuracy (semi-quantitative or qualitative) than required by the current NLLAP accreditation program are not presently covered. This report presents options and issues associated with expanding and redesigning the current NLLAP to cover all laboratories and testing firms generating data for the evaluation of potential lead-poisoning hazards from paint chips, soils, and dust.

1.2 SCOPE OF REPORT

This report develops a field-decision performance based model for NLLAP to address laboratories and lead testing firms that are engaged in at least one of three types of analysis: definitive, semi-quantitative, and qualitative. The model is expressed as a decision tree, with a separate tree being proposed for each type of analysis. Definitive analyses are able to meet strict requirements for accuracy in measurements at the lead action level of concern. Semi-quantitative analyses are defined as quantitative analyses that produce data which do not meet the performance requirements for a definitive analysis, but are still of sufficient quality to support a decision on the presence of hazardous levels of lead in paint chips, dust, and soil with 95 percent confidence. Qualitative analyses do not provide quantitative information about lead levels but are still useful in determining the presence or absence of lead at levels above or below an action level. See Section 1.3 for more complete definitions.

In addition, the performances of three lead testing technologies are evaluated, based upon performance specifications found in literature and past studies. These technologies are portable

x-ray fluorescence (XRF), ultrasonic extraction/anodic stripping voltammetry (UE/ASV), and chemical test kits. The suitability of these instruments for evaluating the presence or absence of lead about an action level is of interest. Specifically, the types of analyses that might be conducted using such technologies are explored. For example, the results of previous studies are considered as evidence about whether or not a laboratory utilizing a given brand of portable XRF technology instruments can be considered a definitive laboratory (i.e., whether definitive-type precision is demonstrated).

1.3 OBJECTIVE OF REPORT

The report is designed to address the following issues:

- Evaluate the lead analysis capability of laboratories and/or testing firms using field portable XRF instruments, UE/ASV instruments, and chemical test kits.
- Identify and define the different types of laboratories or testing firms to be covered under the expanded NLLAP.
- Construct lead analysis decision models for the different types of laboratories or testing firms to be covered under an expanded NLLAP.

1.4 ORGANIZATION OF REPORT

Section 2.0 defines the requirements for a laboratory or testing firm to be categorized as producing definitive data. An analysis of recent ELPAT data for NLLAP accredited laboratories is provided to evaluate the level of performance of those laboratories. The goal of this analysis is to establish the current definition of a definitive laboratory (see Appendix A) as reasonable. A decision rule for evaluating the presence or absence of lead level using definitive analysis is presented.

Section 3.0 defines requirements for a laboratory or testing firm to be categorized as producing semi-quantitative data. The precision of laboratories or testing firms using XRF and UE/ASV technology at lead action levels is evaluated and the requirements for these analyses to be classified as semi-quantitative are discussed. The reason for locating the findings for XRF and UE/ASV technology in this section is because such technologies, as noted later, perform often at the semi-quantitative level. A decision rule for evaluating the presence or absence of lead above an action level using semi-quantitative data is presented.

Section 4.0 defines requirements for a laboratory to be categorized as producing qualitative data. The precision of chemical spot-test kits is evaluated. Chemical test kits are presented in this section because this technology produces qualitative data. A decision rule for evaluating the presence or absence of lead level using qualitative data is presented.

Section 5.0 discusses other important issues related to the expansion of NLLAP, not necessarily covered by the results in this document. The purpose of this section is to raise pertinent unresolved issues in order that they may be properly addressed in the future.

Section 6.0 includes references to the studies evaluated in this report, and other material that is referenced in this report.

A glossary of key terms used in this report is found in Appendix A. Appendix B expands on some of the more detailed statistical issues covered in this report, including evaluating ELPAT data for NLLAP recognized laboratories, the performance of analyses using XRF technology, and the performance of analyses using UE/ASV technology.

2.0 DEFINITIVE LABORATORY/ANALYSIS

This section presents concepts related to analyzing paint, dust, and soil for lead hazard under Definitive Laboratory conditions. Definitions for Definitive Laboratory, Definitive Laboratory gray-zone, and examples are provided.

Every analyzing instrument has some associated measurement error. When a lead level is near the action level, the uncertainty associated with measuring this lead level creates a "gray-zone" in decision making near the action level. For example, if the paint lead level is substantially above (or below) the action level of 1.0 mg/cm^2, then the instrument's uncertainty has minimal impact on the conclusion that the lead level is above (or below) the action level, and it is safe to make that conclusion with a reasonable amount of certainty. If, however, a paint sample is very near 1.0 mg/cm^2, the instrument's imprecision will impact the ability to make the correct conclusion with the desired confidence (e.g., 95 percent confidence), hence creating a "gray-zone."

Laboratories are defined as possessing an inherent gray-zone associated with the instrument or analytical method they use, along with other factors that might affect accuracy. A Definitive Laboratory is defined as a laboratory having a confirmed gray-zone that is no larger than plus or minus 20 percent of an action level with 95 percent confidence. Such a gray-zone is considered to be significantly small.

2.1 DEFINE AND SPECIFY REQUIREMENTS FOR DEFINITIVE CLASSIFICATION

Here, the term "gray-zone" is defined as an approximate 95 percent confidence interval for what the laboratory will observe if the true lead amount is at the action level. Thus an observation within the gray-zone fails to provide sufficient information to allow a conclusion to be made that the true lead amount in the sample is above or below the action level. The definitive requirement that this gray-zone falls within ± 20 percent of the action level roughly translates to a 10 percent coefficient of variation requirement at the action level, to be demonstrated by the laboratory. With the mean and variance of the measurement at the action level represented by μ_A and σ_A^2, respectively, and assuming an unbiased methodology and normality in the measurements, the previous statement can be seen as follows:

$$95\%\ \text{C.I.} \Rightarrow \mu_A \pm 2\sigma_A\,,$$
$$20\%\ \text{rule} \Rightarrow (2\sigma_A)/\mu_A \leq 0.20\,,$$

so (σ_A/μ_A), the coefficient of variation at the action level, must be ≤ 0.10, or 10 percent.

When action levels correspond to the proposed §403 hazard standards, the following list translates the above-defined Definitive Laboratory requirement to gray-zone requirements for analysis of paint, dust, and soil samples:

- 1.0 ± 0.2 mg/cm^2 for paint (or 0.5 % ± 0.1 % lead by weight) (under the statutory definitions for lead-based paint)
- 50 ± 10 $\mu g/ft^2$ for dust on floors

- 250 ± 50 μg/ft^2 for dust on window sills
- 800 ± 160 μg/ft^2 for dust on window wells (under the interim §403 guidance)
- 2,000 ± 400 ppm for soil.

2.2 ANALYSIS OF ELPAT DATA FOR NLLAP RECOGNIZED LABORATORIES

ELPAT data from rounds 14 to 21 for NLLAP recognized laboratories were considered for analysis in order to determine whether such laboratories, currently considered definitive in performance, do in fact achieve such precision. For each medium (paint, dust, and soils), each lab analyzed four samples per round, yielding a total of 32 analyses per lab. The overall set of observed mean values were treated as true values, and the corresponding standard errors were treated as a function of the truth. Ordinary least squares regression was used to determine the parameters defining the approximate linear relationship between the lead level and the precision of NLLAP recognized laboratories. Details about the statistical modeling used in this analysis, as well as summary statistics for available data, are provided in Appendix B, Section 1. Note that the parameters presented in Table B.1 are used to calculate a "precision" value at the action levels of concern. This precision is then used to calculate the gray-zone, which equals the action level plus or minus 2 times the precision. The theoretical Definitive Laboratory gray-zones and the NLLAP observed gray-zones for paint, dust, and soils are presented below in Table 2.1.

Table 2.1 Gray-zone comparison between definitive laboratory requirements and NLLAP recognized laboratories

Medium		Action Level (± 20%)	Definitive Gray-Zone	NLLAP Observed Gray-Zone
Paint (mg/cm^2)		1.0 ± 0.2	[0.80, 1.20]	[0.87, 1.13]
Dust (ug/ft^2)	Floor	50 ± 10	[40.0, 60.0]	[36.13, 63.87][1]
	Window Sills	250 ± 50	[200, 300]	[210, 290]
	Window Wells	800 ± 160	[640, 960]	[688, 910]
Soil (ppm)		2000 ± 400	[1600, 2400]	[1789, 2211]

[1] Based on data mostly above 50 ug/ft^2 (see Figure B.1 of Appendix B). Result should be interpreted with caution.

The results show that on average, the NLLAP recognized laboratories are able to perform within the prescribed standard of the action level plus or minus 20 percent. The definitive-type performance of the NLLAP recognized laboratories follows from the fact that the NLLAP observed gray-zones all are narrower than the Definitive gray-zone requirements in Table 2.1, except for floor dust.

2.3 PRESENT AND DEFINE A DECISION TREE FOR 95 PERCENT CONFIDENCE

Figure 2.1 below presents a decision tree for making a decision with approximate 95 percent confidence that a true lead level is above or below an action level, when using a Definitive Laboratory to perform the analysis. Notice that even though a laboratory may be considered definitive, it still has an associated amount of imprecision, as recognized by its

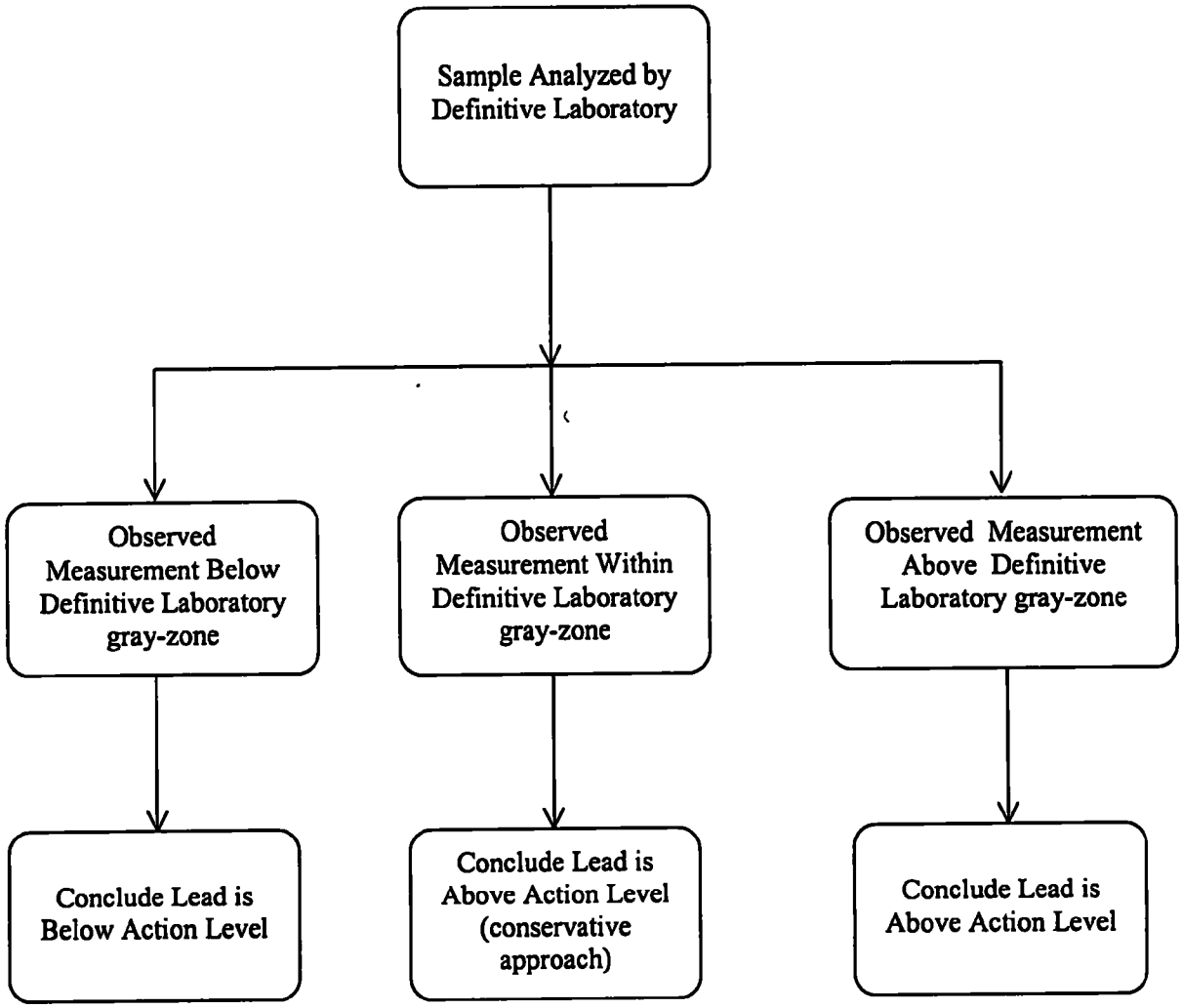


Figure 2.1 Decision tree for making statements with 95 percent confidence using a Definitive Laboratory.

gray-zone. As such, results in the gray-zone of a Definitive Laboratory cannot be classified with 95 percent confidence as either above or below the action level. However, if some decision must be made, and since a gray-zone result does not provide clear evidence that the true lead amount is below the action level, then the decision tree is conservative by concluding that lead is above the action level in such cases. The net result of this choice is that an increased frequency of false positive classifications will occur, associated with lead levels that are truly below an action level but classified otherwise due to instrument imprecision. However, such a misclassification is preferred to a false negative classification. That is, the conservative approach of increasing false positive classifications protects those at risk for lead exposure.

3.0 SEMI-QUANTITATIVE LABORATORY/ANALYSIS

A Semi-Quantitative Laboratory is a laboratory performing a quantitative analysis, but whose methods provide a gray-zone (i.e., 95 percent confidence interval) that is wider than plus or minus 20 percent of the action level. The decision tree to be developed below recommends that any sample whose lead measurement falls within a Semi-Quantitative Laboratory gray-zone be sent to a Definitive Laboratory for more accurate analysis. The results from the Definitive Laboratory then will be used to make the final conclusion with respect to the action level.

As an example, consider a laboratory using an LPA-1 portable XRF as its measurement technology. According to the performance characteristic sheet (PCS) of the LPA-1, with a reading measured at 1.0 mg/cm^2 for lead in paint, the substrates brick, concrete, drywall, plaster, and wood have zero bias and a precision of 0.3 mg/cm^2. Therefore, LPA-1 instruments have a gray-zone of $1.0 \pm 2*0.3$, or (0.4, 1.6) mg/cm^2, for the listed substrates. Since the LPA-1 gray-zone is wider than the Definitive Laboratory gray-zone of (0.8, 1.2) mg/cm^2, the PCS numbers suggest that a laboratory using the LPA-1 XRF analyzer would be classified as a Semi-Quantitative Laboratory for analyzing lead in paint.

As this example suggests, it is expected that data for portable XRF and UE/ASV technologies will suggest that laboratories using such techniques often will be classified as Semi-Quantitative Laboratories. Therefore, findings for portable XRF and UE/ASV technologies are presented in this section. However, it is not necessarily the case that such technologies always lack the precision to be considered definitive in nature, as the discussion below will show.

3.1 CHARACTERIZE XRF PRECISION

This section considers the performance of portable XRF technologies. In section 3.1.1, the gray-zones as determined from PCS information are presented. A 1993 field study, described in more detail in section 3.1.2, uses portable XRF instruments in the field, following manufacturer instructions. An additional field study performed by the National Institute for Occupational Safety and Health (NIOSH) to evaluate three lead-based paint detection technologies is discussed in Section 3.1.3. Finally, analysis of lead in soil by portable XRF is considered in Section 3.1.4.

3.1.1 XRF Precision from Performance Characteristic Sheets

This section presents estimated gray-zones for various makes and models of portable XRF instruments, according to data provided by XRF Performance Characteristic Sheets (PCSs). First, some background is provided. In (EPA 747-R-95-008), bias and precision estimates for XRF instruments are obtained from field testing data. Since estimates of bias and precision are based on analysis of field samples, the report considers two fundamental issues:

1. The lead levels within the field samples were more distributed toward lower values, with fewer samples occurring as the level increases.

2. Lead levels are themselves estimated by laboratory analysis of paint samples using inductively coupled plasma-atomic emission spectroscopy (ICP-AES), which itself has some measurement error.

These two factors make it impossible to directly observe the bias and precision of XRF results under field conditions, or at pre-specified lead levels. In order to estimate XRF bias and precision, the report makes several assumptions, which are listed below:

- A linear regression relationship exists between the mean XRF measurements and the true lead level, i.e., XRF = a + b*(true lead level) + error.
- The magnitude of the regression error is proportional to true lead level. The model used is Error = c + d*(true lead level).
- It was assumed that the distribution of true lead levels is lognormal, and that ICP-AES measures the natural logarithm of the true lead level with a known measurement error. This assumption affects how the parameters (a, b, c, and d) are estimated.

For further details on deriving the estimates to be used in a PCS, see (EPA 747-R-95-008).

An XRF PCS is instrument-model specific and is created to provide testing guidance and detailed performance information. This information includes the specification of conclusive and inconclusive XRF results. The PCS also provides calibration check values to be used in conjunction with the NIST Standard Reference Material paint films and a procedure for evaluating XRF testing. Table 3.1 presents eight XRF manufacturer's make and model information, and gray-zones based on precision and bias reported in the PCS. Performance is stratified by substrate. The results of Table 3.1 provide evidence that a laboratory utilizing a portable XRF instrument for analysis would probably be classified as a Semi-Quantitative Laboratory. That is, the observed gray-zones in Table 3.1 are wider than the definitive requirement of (0.8, 1.2) mg/cm^2 for lead in paint.

Table 3.1 Gray-zones (mg/cm^2) for various XRF instruments when measuring lead levels on painted surfaces, based on PCS data

Manufacturer	Make and Model	Measured at	Brick	Concrete	Drywall	Metal	Plaster	Wood
TN Technologies	Pb Analyzer 9292	Normal reading time at 15-seconds	(0.2, 1.8)	(0.2, 1.8)	(0.2, 1.8)	(0.2, 1.8)	(0.2, 1.8)	(0.2, 1.8)
Scitec Corporation	MAP-3	15-second reading	(0.0, 3.0)	(0.0, 3.0)	(0.2, 1.8)	(0.0, 2.2)	(0.0, 2.8)	(0.0, 2.4)
		60-second reading	(0.0, 2.4)	(0.0, 2.4)	(0.4, 1.6)	(0.2, 1.8)	(0.0, 2.6)	(0.2, 1.8)
Warrington, Inc.	Microlead I revision 4	Normal reading time at 15-seconds	(0.0, 2.2)	(0.0, 2.4)	(0.4, 1.6)	(0.0, 2.2)	(0.0, 2.4)	(0.0, 2.4)
Princeton Gamma-Tech, Inc.	XK-3	Normal reading time at 15-seconds	(0.0, 2.2)	(0.0, 2.4)	(0.2, 1.8)	(0.0, 3.0)	(0.0, 2.2)	(0.0, 2.4)
Niton Corporation	XL-309, 701-A, 703-A Spectrum Analyzers	Variable time mode, software version 5.1	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)
Radiation Monitoring Devices	LPA-1 sold prior to /serviced before June 26, 1995	20-second reading	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)
		30-second reading	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)
		Quick Mode	(0.2, 1.8)	(0.2, 1.8)	(0.2, 1.8)	(0.2, 1.8)	(0.2, 1.8)	(0.2, 1.8)
	LPA-1 sold or serviced after June 26, 1995	20-second reading	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)
		Quick Mode	(0.2, 1.8)	(0.2, 1.8)	(0.2, 1.8)	(0.2, 1.8)	(0.2, 1.8)	(0.2, 1.8)
Scitec Corporation	MAP-4	Screen Mode	(0.2, 1.8)	(0.2, 1.8)	(0.0, 2.2)	(0.4, 1.6)	(0.2, 1.8)	(0.0, 2.2)
		Test Mode	(0.4, 1.6)	(0.4, 1.6)	(0.0, 2.2)	(0.6, 1.4)	(0.4, 1.6)	(0.0, 2.2)
Advanced Detectors, Inc.	LeadStar w/ software v 4.1 to 4.30	Fixed Mode	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)
	LeadStar w/ software versions less than 4.1	Fixed Mode	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)	(0.4, 1.6)

Note: Gray-zones are defined as 1.0 mg/cm^2, ± two times the documented precision. The definitive gray-zone of comparison is (0.8, 1.2) mg/cm^2.

3.1.2 EPA Field Study on XRF Measurement Precision

In 1993, a study was conducted by the U.S. EPA (EPA 747/R-95/002b) to collect information necessary to establish federal guidelines on testing for lead in paint. The overall study objective was "to collect information about field measurement methodologies sufficient to allow EPA and HUD to establish guidance and protocols for lead hazard identification and evaluation." Included in this report is a statistical model that was used to describe the relationship between XRF measurements and the lead level as analyzed by ICP-AES (see Appendix B, Section 2, for more detail). ICP is a method commonly used in laboratories to analyze lead in paint and is one of the techniques recommended for confirmation testing.

For each of the six field portable devices tested in this study (Lead Analyzer, MAP-3, Microlead I, X-MET 880, XK-3, and XL), instrument gray-zones were calculated using a precision estimate derived from the statistical model parameters provided in the report. Since the Lead Analyzer and MAP-3 instruments could be operated by using either K-shell or L-shell X-rays, results were recorded once for each energy level. Table 3.2 presents the gray-zone values for XRF instruments based on this field study. As in the previous section, the results provide evidence that a laboratory using a portable XRF instrument for analysis would likely be classified as a Semi-Quantitative Laboratory.

Table 3.2 Gray-zones defined as 1.0 ± two times the precision in mg/cm^2, where 0 is the lowest limit. Results are based on the above described field test data.

Instrument	Energy Level	Brick	Concrete	Drywall	Metal	Plaster	Wood
Lead Analyzer	K-shell	(0.527, 1.473)	(0.26, 1.74)	(0.29, 1.71)	(0.185, 1.815)	(0.527, 1.473)	(0.135, 1.865)
	L-shell	(0.925, 1.075)	(0.799, 1.201)	(0.659, 1.341)	(0.641, 1.359)	(0.724, 1.276)	(0.63, 1.37)
MAP-3	K-shell	(0.00, 2.864)	(0.00, 2.989)	*N/A or (0.249, 1.751)	(0.00, 2.094)	(0.00, 2.733)	(0.00, 2.33)
	L-shell	(0.518, 1.482)	(0.671, 1.329)	(0.506, 1.494)	(0.086, 1.914)	(0.697, 1.303)	(0.449, 1.551)
Microlead I	K-shell	N/A	(0.00, 2.439)	*N/A or (0.322, 1.678)	(0.00, 2.368)	(0.00, 2.238)	(0.00, 2.83)
X-MET 880	L-shell	N/A	(0.859, 1.141)	(0.73, 1.27)	(0.523, 1.477)	(0.839, 1.161)	(0.422, 1.578)
XK-3	K-shell	*N/A or (0.00, 2.198)	*N/A or (0.00, 2.271)	(0.00, 2.124)	(0.00, 3.116)	(0.00, 2.266)	(0.00, 2.373)
XL	L-shell	(0.631, 1.369)	(0.514, 1.486)	(0.527, 1.473)	(0.22, 1.78)	(0.671, 1.329)	(0.371, 1.629)

* Calculated with d = 0, could be entered as N/A or as gray-zone given. (See Appendix B)
N/A = Not applicable (insufficient data)
Note: The definitive gray-zone of comparison is (0.8, 1.2) mg/cm^2.

3.1.3 Field Investigation of On-Site Techniques (Portable XRF)

Three field lead detection technologies for detecting lead levels in paint were evaluated in (Ashley, et al., 1998-2), one of which was XRF. This was a field study, conducted by NIOSH on buildings erected from the late 1800's to the 1960's on the campus of Florida A&M University in Tallahassee. The XRF field instrument used was the TN Spectrace 2000. Confirmatory analyses of split paint samples were carried out by an accredited laboratory using atomic absorption spectrometry (AAS), and the intent of the study was primarily to determine the level of bias in on-site measurements.

A total of 175 measurements were taken on paint test readings for various media (plaster, metal, wood, and brick), and the XRF test results were compared to AAS results by linear regression. The mean-squared error for the regression was given, as well as the slope estimate, intercept estimate, and r-square value. The action level for paint of 0.5 percent lead by weight was lower than the average paint concentration observed, and variability tended to increase with concentration. Because of this, it is assumed that the reported standard error, which is the root mean-squared error of the regression line, is probably larger than error associated with lead levels close to the action level. The reported standard error is 0.054 percent, resulting in a gray-zone of (0.392-0.608) percent around the action level, and a coefficient of variation of 0.108. Since the definitive gray-zone requirement in this case is (0.4, 0.6) percent, the observed performance using this XRF technology does not quite meet definitive-type standards.

It should be noted that the estimate of standard error given here is not completely accurate at the action level, although it is reasonable to assume the true standard error is lower. From this perspective, the result is conservative. Also, results from the XRF instruments appear to be biased high when compared to reference concentrations determined by AAS. XRF readings gave results that were, on average, higher than the reference concentrations. Of course, if the bias is well understood, it can be corrected for.

3.1.4 Using XRF Technology for Soil Analysis

The on-site capabilities of XRF instrumentation for measuring lead in soil was investigated in (EPA 600-R-97-145). Two instruments were used in this analysis: the Spectracle TN Pb Analyzer and the Spectracle TN 9000 Analyzer. Two sites were selected to perform this on-site analysis: one in Maryland and a second in Iowa. Heavy industrial activity had taken place at both sites. Samples collected on-site were split and evaluated by a reference laboratory using ICP-AES to provide a reference concentration.

Soil samples were classified into groups according to their reference lead concentrations. Then ten replicate measurements for lead were made on the soil samples using the same instrument by the same individual. The results here are, therefore, estimates of repeatability of measurements, and not reproducibility.

The lead concentrations of soils were reported only as falling into one of four categories: (1) near the minimum detection level, (2) 50-500 μg lead/gram soil, (3) 500-1,000 μg/g, and (4) >1,000 μg/g. Twenty soil samples were found to contain reference concentrations of lead that were >1,000 μg/g, which is the category containing the 2,000 μg/g action level for soils. Results here were reported as relative standard deviations (RSD), which is the ratio of standard deviation to the category mean. The TN Pb Analyzer was reported as having a RSD of 2.52 for lead, which would give a gray-zone of (1,900, 2,100) μg/g around the action level, as compared to the definitive requirement of (1,600, 2,400) μg/g. The TN 9000 Analyzer was reported as having a RSD of 3.68 for lead, which yields a gray-zone of (1,853, 2,147) μg/g around the action level. In both cases, the performances appear to be of definitive-type quality for analysis of lead in soil. However, it should be noted that the only source of variability estimated in this study was associated with repeated measurements of the same sample by the same instrument and individual, and any additional variation due to reproducibility is not taken into account.

3.2 CHARACTERIZING ULTRASONIC EXTRACTION/ANODIC STRIPPING VOLTAMMETRY (UE/ASV) PRECISION

This section considers the performance of Ultrasonic Extraction/Anodic Stripping Voltammetry (UE/ASV) technologies. Overall, UE/ASV roughly demonstrates a level of precision and accuracy that is potentially compatible with definitive-type analysis. However, as seen below, this does not appear to always be the case.

3.2.1 EPA Evaluation of the PaceScan 2000

A study was performed by the U.S. EPA (EPA 600/R-95/093) to evaluate the performance of solution-based technologies for measuring lead in environmental media. One of the instruments tested was the PaceScan 2000, which uses UE/ASV technology. Accuracy and precision of the instrument were determined by taking measurements of lead in characterized paints, bulk dusts, and soils, designated as Research Triangle Institute (RTI) core materials. These RTI core materials include materials from the following sources:

- NIST Standard Reference Materials (NIST SRMs) – reference samples prepared and certified by the National Institute of Standards and Technology.

- RTI Method Evaluation Materials (RTI MEMs) – samples with lead concentrations determined from an EPA/RTI round-robin study (Williams, et al., 1993) done by using hotplate or microwave extraction, with measurement by AAS or ICP-AES.

- ELPAT Materials – reference laboratory samples with mean concentrations determined by a number of reference laboratories selected by NIOSH using a range of extraction methods with measurement by AAS or ICP-AES.

For the purpose of analysis, it was assumed that the measured mean lead concentrations of these reference materials were in fact the true lead concentration. Three repeated measurements were made on a number of paint, bulk dust, and soil samples. It was also assumed

that the instrument would have similar characteristics for bulk dust and dust wipes, as bulk dust was analyzed in the study.

Because the data were presented in full and are of high quality, a statistical analysis was performed on the data to identify precision at the specified action levels. Details of this analysis are provided in Appendix B. The results presented in Table 3.3 indicate that definitive analysis criteria are satisfied for paint and soil testing, but only semi-quantitative analysis criteria were met for bulk dust testing.

The PaceScan 2000 has both a high and a low setting for paint: The low range is 0.0025-1.5 percent lead by weight, and the high range is 0.02-10 percent. The lead detection range for dust and soil samples is 0.0025-1.5 percent. The instrument also has extended ranges that were not considered. The report also concluded that "The PaceScan 2000 instrument provided applicability to multimedia analysis, was easily operated, and appeared to have promise for field application".

Table 3.3 PaceScan 2000 results based on data from (EPA 600/R-95/093), April 1996

Medium	Action Level	Precision	Percent Error	Gray-Zone	Definitive Zone	Within 20%
Paint – Low	0.5%	0.032%	6.4%	0.436-0.564	0.400-0.600	Y
Paint – High	0.5%	0.042%	8.3%	0.417-0.583	0.400-0.600	Y
Bulk Dust	50 µg/sample	6.2 µg/sample	12.4%	38-62	40-60	N
Bulk Dust	250 µg/sample	27 µg/sample	10.8%	196-304	200-300	N
Bulk Dust	800 µg/sample	86 µg/sample	10.7%	628-972	640-960	N
Soil	2000 µg/g	130 µg/g	6.5%	1740-2260	1600-3200	Y

Dust samples are assumed to represent a 1 ft^2 area.

3.2.2 Interlaboratory Evaluation of UE/ASV Lead Measurements on Paint, Dust, and Soil

An interlaboratory evaluation of the UE/ASV procedure obtaining estimates for both repeatability and reproducibility of measurements was conducted by NIOSH and reported in (Ashley, et al., 1998-1). The UE/ASV technology evaluated was the Palintest 5000, a later version of the PaceScan 2000. Paint, soil, and dust wipe performance evaluation materials (PEM's) prepared by the Research Triangle Institute (RTI) were used in the study. These samples were collected at various commercial and residential sites in several states, and then dried, ground, sieved, and homogenized prior to initial laboratory characterizations for lead content using UE in accordance with ASTM PS87 and ASV in accordance with ASTM PS88. Dusts were spiked onto wipes. As a reference analytical method, ICP-AES with microwave digestion was used to characterize the samples.

Measurements performed on the RTI reference samples were chosen to bracket action levels for the different matrices. Two measurements were taken at each of ten laboratories, at three different lead concentration levels per medium. Due to a procedural error at one of the laboratories, the paper contains results from only nine laboratories.

In order to address questions about precision of UE/ASV measurements at action levels, a few general assumptions were made about the data. These assumptions are detailed in Appendix B. Results are shown in Table 3.4, which again demonstrates a mixture of definitive and semi-quantitative performance for analyses using UE/ASV technology.

Table 3.4 Results based on data from interlaboratory evaluation of UE/ASV (Palintest 5000)

Medium	Action Level	Precision	Percent Error	Gray-Zone	Definitive Zone	Within 20%
Paint	0.5%	0.070%	14.0%	0.360-0.640	0.400-0.600	N
Dust – floors	50 µg/sample	6 µg/sample	12.4%	38-62	40-60	N
Dust-sills	250 µg/sample	24 µg/sample	9.6%	202-298	200-300	Y
Dust-wells	800 µg/sample	53 µg/sample	6.7%	694-906	640-960	Y
Soil	2000 µg/g	136 µg/g	6.8%	1728-2272	1600-2400	Y

3.2.3 Field Investigation of On-Site Techniques (UE/ASV)

Three field lead detection technologies for detecting lead levels in paint, including UE/ASV, were evaluated in the NIOSH study presented in Section 3.1.3 (Ashley, et al., 1998-2). The UE/ASV field instrument used was the PaceScan 3000. Confirmatory analyses of split paint samples were carried out by an accredited laboratory using AAS, and the intent of the study was primarily to determine the level of bias in on-site measurements.

A total of 71 analyses were taken on paint test samples from various media (plaster, metal, wood, and brick), and the UE/ASV test results were compared to AAS results by linear regression. The mean-squared error for the regression was given, as well as the slope estimate, intercept estimate, and r-square value. The action level for paint of 0.5 percent was lower than the average paint lead concentration investigated, and variability tends to increase with concentration. Because of this it is assumed that the reported standard error, which is the mean squared error of the regression line, is probably larger than error at the action level. Reported standard error is 0.0390 percent, resulting in a gray-zone of 0.4220-0.5780 percent around the action level, and a coefficient of variation of 0.078. This gives a gray-zone that has all values within 20 percent of the action level, indicating definitive-type performance.

It should be noted that the estimate of standard error given here is not completely accurate at the action level, although it is reasonable to assume the true standard error is lower. From this perspective, the result is conservative. Also, two outliers were deleted before analysis of the data. However, these outliers occurred at levels much higher than the action level and did not significantly impact estimation of precision.

3.2.4 Laboratory Evaluation of the PaceScan 2000

The effectiveness of the PaceScan 2000 in performing analyses of real-world waste toxic characteristic leaching procedure (TCLP) samples was evaluated in (White and Clapp, 1998). TCLP samples are samples taken from remediation sites and are not classified as paint, dust, or soil samples but the dust setting was used to evaluate the lead concentration of the samples.

Analysis results from ICP-AES done in extract were used as a reference. The comparative analysis was done on leachate, the solution that was extracted.

Eighteen TCLP samples were analyzed using both ICP-AES analysis and using the PaceScan 2000 with sample aliquots acidified to 2 percent and 4 percent concentrated nitric acid solutions. The study noted that there was no significant difference between the 2 and 4 percent acidification results and also no significant difference between either of these results and the ICP-AES reference readings.

A spike recovery study was performed using spikes of known amounts of lead followed by repeated analysis of the spike using the PaceScan 2000. The coefficient of variation of these measurements never exceeded 3 percent. This study showed that the dust setting of the PaceScan 2000 can give results that could be qualified as a definitive-type analysis under controlled laboratory conditions.

A repeatability study was also done by performing ten parallel analyses of three TCLP sample extracts which had concentrations near lead action levels for dust. The reference concentrations from the ICP-AES analysis were 80, 550, and 650 µg/sample. Using linear extrapolation to estimate variance of measurements at dust action levels is difficult for this study as the observed standard deviation at the 550 µg/sample level was actually larger than the observed standard deviation for 650 µg/sample. To estimate the standard deviation at 50 µg/sample, the observed standard deviation at 80 µg/sample was used. For the 250 and 800 µg/sample action levels, standard deviations were estimated by a linear regression equation. Results are summarized in Table 3.5 and suggest definitive-type performance.

Table 3.5 Results based on data from analysis of TCLP Extracts (PaceScan 2000)

Medium	Action Level	Precision	Percent Error	Gray-Zone	Definitive Zone	Within 20%
Dust – floors	50 µg/sample	2.4 µg/sample	4.8%	45-55	40-50	Y
Dust-sills	250 µg/sample	17 µg/sample	6.8%	216-284	200-300	Y
Dust-wells	800 µg/sample	54 µg/sample	6.7%	694-908	640-960	Y

3.2.5 Other Studies Considered

One study (Ashley, 1995) investigated UE/ASV technology but performed analysis on air filter samples. The study concluded that UE/ASV technology gave definitive-type results in the neighborhood of action levels for dust wipes, which is considered a comparable analysis.

Finally, manufacturer claims for the Palintest 5000 are that the scanner has a coefficient of variation of less than or equal to 7 percent at the action levels for paint, dust wipes, and soil. If appropriate, these results, along with the results in the previous sections, indicate that UE/ASV technology may possess the potential for being used in a definitive-type analysis.

3.3 PRESENT AND DEFINE A DECISION TREE FOR 95 PERCENT CONFIDENCE

Recall that the gray-zone for an instrument (action level, plus or minus two times the known precision for that instrument) represents an approximate 95 percent confidence interval for what the instrument will observe if the true lead amount is at the action level. Thus an observation within the gray-zone fails to provide strong enough evidence regarding whether the true lead amount is above or below the action level.

Recall that a Definitive Laboratory has a confirmed gray-zone at least as narrow as plus or minus 20 percent of the action level, while a Semi-Quantitative Laboratory is a laboratory whose methods provide a gray-zone that is wider than plus or minus 20 percent of the action level. Figure 3.1 below provides a decision tree for providing 95 percent confidence in making decisions relative to an action level when using a Semi-Quantitative Laboratory to perform the analysis. The tree indicates that initial gray-zone results should be sent to a Definitive Laboratory for confirmation. The idea is that the imprecision associated with a Semi-Quantitative Laboratory will produce a large number of gray-zone results. Therefore, in order to avoid a high rate of false positive or false negative classifications due to making a decision based on gray-zone results, such results require the more precise analysis of a Definitive Laboratory.

The following scenarios correspond to a Semi-Quantitative Laboratory using portable XRF technology with performance similar to that given on the LPA-1 PCS. These scenarios are meant to illustrate the use of the decision tree given by Figure 3.1.

- The observed paint lead measurement is 0.2 mg/cm^2. The observed value is below the lower limit of the Semi-Quantitative gray-zone (0.4, 1.6). The appropriate conclusion is that the lead loading is below the action level of 1.0 mg/cm^2. This conclusion is made with 95 percent confidence.
- The observed paint lead measurement is 2.6 mg/cm^2. The observed value is above the upper limit of the Semi-Quantitative gray-zone (0.4, 1.6). The appropriate conclusion is that the lead loading is above the action level of 1.0 mg/cm^2. This conclusion is made with 95 percent confidence.
- The observed paint lead measurement is 0.8 mg/cm^2. The observed value is within the Semi-Quantitative gray-zone (0.4, 1.6). The appropriate action is to send the sample to a Definitive Laboratory (gray-zone within (0.8,1.2)). The Definitive Laboratory will analyze the sample. The appropriate conclusion is made based on the results from the Definitive Laboratory.

In the last scenario presented above, observe that the possibility exists, due to Definitive Laboratory imprecision, that the subsequent result will lie within a gray-zone as well. As such, if a decision must be made at this point, then it is not necessarily made with 95 percent confidence. Essentially, once an initial gray-zone result is obtained by the Semi-Quantitative Laboratory, the decision tree in Figure 3.1 defaults to the decision tree in Figure 2.1.

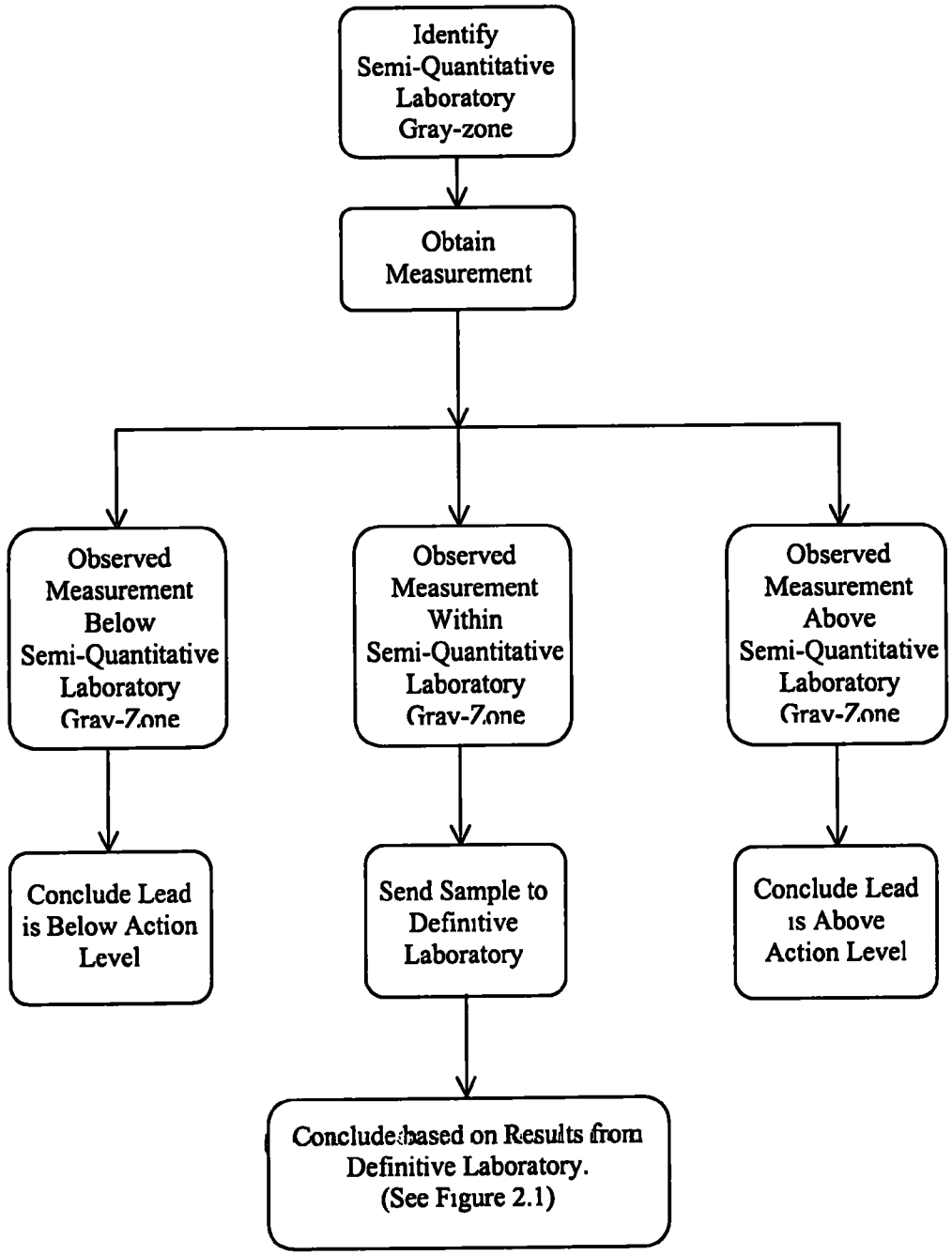


Figure 3.1 Decision tree for making statements with 95 percent confidence using a Semi-Quantitative Laboratory.

4.0 QUALITATIVE LABORATORY/ANALYSIS

This section discusses the use of qualitative measures for lead in paint, dust, or soil. Specifically, chemical test kits that provide only an indication of presence or absence of lead are considered. The issue of whether such qualitative analyses can be used for the purpose of making a decision, at 95 percent confidence, with respect to a true lead level as compared to an action level is investigated.

The general finding is that qualitative analyses, such as that performed in the application of a chemical test kit, may be appropriate for making a decision, at 95 percent confidence, in a single direction but probably not in two directions. Specifically, Section 4.1 discusses using chemical test kits as a negative screen (i.e., concluding that the amount of lead is below the action level). Section 4.2 discusses using chemical test kits as a positive screen (i.e., concluding that the amount of lead is above the action level). Section 4.3 discusses some field results for different chemical test kits. Finally, Section 4.4 presents a decision-tree model for combining information from the two types of qualitative measures discussed in Sections 4.1 and 4.2.

4.1 USING QUALITATIVE ANALYSES AS NEGATIVE SCREENS

Figure 4.1 demonstrates hypothetical performance of a chemical test kit for the analysis of lead in paint. This figure is an operating characteristic (OC) curve which plots the probability of a positive indication for lead in paint as a function of the true amount of lead in paint. The vertical dashed line corresponds to the action level of 1.0 mg/cm^2 for lead in paint. The horizontal dashed line corresponds to a 95 percent probability of obtaining a positive indication for lead in paint.

The hypothetical results demonstrated in Figure 4.1 represent the type of qualitative performance that would be appropriate for making decisions in the direction of a negative screen. The figure shows that if the true lead level is at or above the action level of 1.0 mg/cm^2, then at least 95 percent of the time a positive indication will be obtained:

$$\text{Probability } \{\text{Positive indication given true lead level} \geq 1.0 \text{ mg/cm}^2\} \geq 0.95.$$

This implies that the test has high sensitivity. Equivalently, the likelihood of a false negative (i.e., having a negative indication when the time lead level is at or above the action level), is no more that 5 percent:

$$\text{Probability } \{\text{Negative indication given true lead level} \geq 1.0 \text{ mg/cm}^2\} \leq 0.05.$$

Such performance provides 95 percent confidence that negative indications, where made, are correct indications. Therefore, a chemical test kit exhibiting this type of performance could be used as a negative screen.

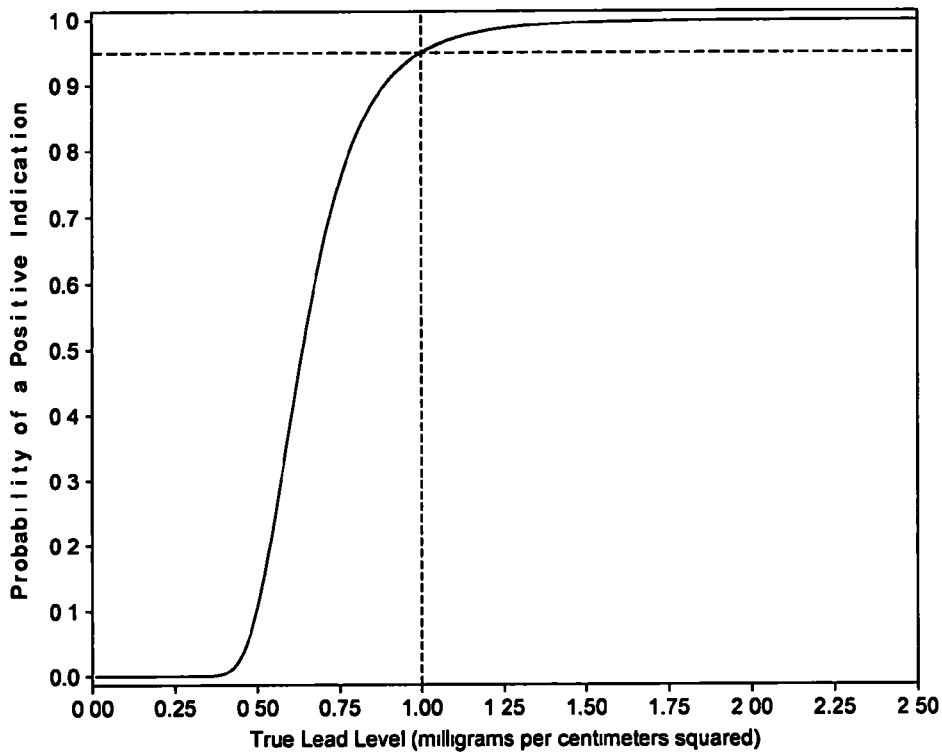


Figure 4.1 Hypothetical operating characteristic (OC) curve of a chemical test kit analyzing lead in paint (demonstrating qualitative analysis performance considered appropriate as a negative screen).

In contrast to the above discussion, Figure 4.1 also highlights the fact that a chemical test kit exhibiting the displayed performance would not necessarily be appropriate for making decisions based on positive indications. The high sensitivity (i.e., likelihood of a positive result) of such an instrument at lower lead levels would produce far too many false positive results. Specifically, for lead levels between 0.5 and 1.0 mg/cm^2, this test kit would provide an incorrect positive indication more than 5 percent of the time. Such results are considered false positives given the action level of 1.0 mg/cm^2. For this reason, a more appropriate course of action given a positive indication by this test kit would be to send a sample to a definitive laboratory in order to obtain a more accurate quantitative result.

In summary, a qualitative measurement technology such as a chemical test kit may have potential for assessing lead in paint, dust, or soil – with 95 percent confidence. Unfortunately, the qualitative nature of the analysis being performed appears to produce a limitation on the decisions that can be drawn from obtained results. However, a straightforward decision tree that provides an overall protection of 95 percent confidence against incorrect conclusions still can be formed. In the case of a qualitative analysis used as a negative screen, the decision tree providing 95 percent confidence against error would be as follows:

- Perform the analysis and obtain a result.
  (a) If the result is a negative indication for lead, conclude with 95 percent confidence that the true lead level is below the action level.
  (b) If the result is a positive indication for lead, send a sample to a definitive laboratory for quantitative confirmation with 95 percent confidence.

4.2 USING QUALITATIVE ANALYSES AS POSITIVE SCREENS

Figure 4.2 presents an alternative OC curve, where the horizontal dashed line corresponds to a 5 percent probability of obtaining a positive indication for lead in paint. The hypothetical results demonstrated in Figure 4.2 represent the type of qualitative performance that would be appropriate for making decisions in the direction of a positive screen. In other words, a decision could be made, with 95 percent confidence, as to whether the amount of lead is above the action level. The figure shows that if the true lead level is below the action level of 1.0 mg/cm^2, then at least 95 percent of the time a negative indication will be obtained:

$$\text{Probability } \{\text{Negative indication given true lead level} < 1.0 \text{ mg/cm}^2\} \geq 0.95.$$

This implies that the test has high specificity. Equivalently, the likelihood of a false positive (i.e., having a positive indication when the time lead level is below the action level) is no more than 5 percent:

$$\text{Probability } \{\text{Positive indication given true lead level} < 1.0 \text{ mg/cm}^2\} \leq 0.05.$$

Such performance provides 95 percent confidence that positive indications, where made, are correct indications. Therefore, a chemical test kit exhibiting this type of performance could be used as a positive screen.

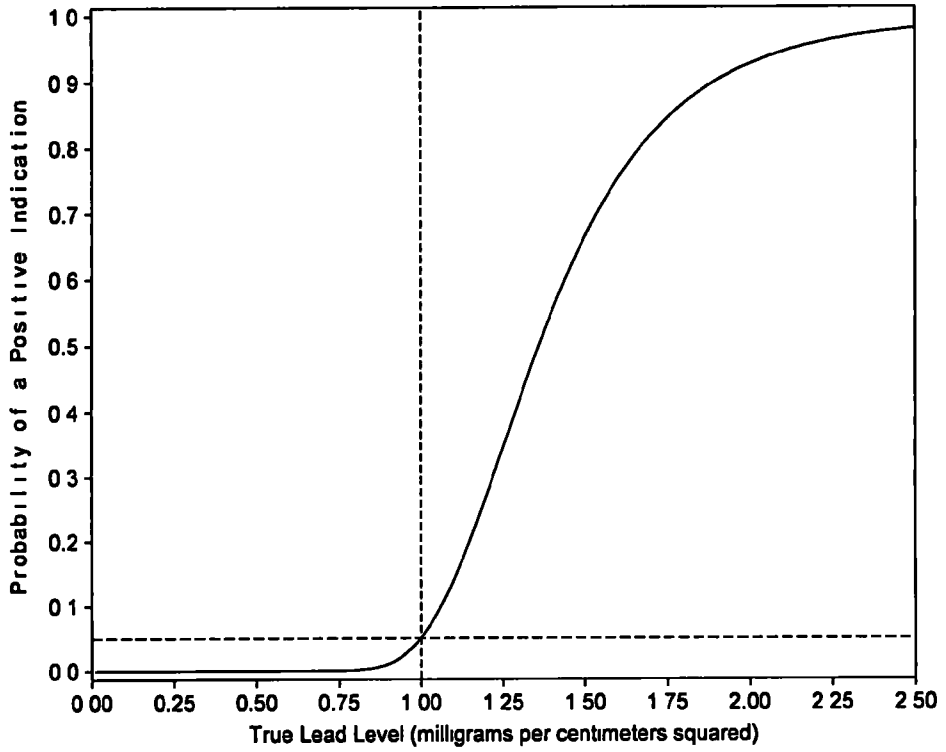


Figure 4.2 Hypothetical operating characteristic (OC) curve of a chemical test kit analyzing lead in paint (demonstrating qualitative analysis performance considered appropriate as a positive screen).

Figure 4.2 also highlights the fact that a chemical test kit exhibiting the displayed performance would not necessarily be appropriate for making decisions based on negative indications. The high specificity (i.e., likelihood of a negative result) of such an instrument at lead levels near 1.0 mg/cm^2 would produce far too many false negative results. Specifically, for lead levels between 1.0 and 2.0 mg/cm^2, this test kit would provide an incorrect negative indication more than 5 percent of the time. Such results are considered false negatives given the action level of 1.0 mg/cm^2. For this reason, a more appropriate course of action given a negative indication by this test kit would be to send a sample to a definitive laboratory in order to obtain a more accurate quantitative result.

As seen in Section 4.1, the qualitative nature of the analysis being performed appears to produce a limitation on the decisions that can be drawn from obtained results. Again however, a straightforward decision tree that provides an overall protection of 95 percent confidence against incorrect conclusions can be formed. In the case of a qualitative analysis used as a positive screen, the decision tree providing 95 percent confidence against error would be as follows:

- Perform the analysis and obtain a result.
  (a) If the result is a positive indication for lead, conclude with 95 percent confidence that the true lead level is above the action level.
  (b) If the result is a negative indication for lead, send a sample to a definitive laboratory for quantitative confirmation with 95 percent confidence.

4.3 FIELD TEST RESULTS FROM EVALUATING CHEMICAL TEST KITS

This section discusses two different studies that considered the performance of chemical test kits. The results of these two field studies are given below.

4.3.1 EPA Field Study on Chemical Test Kits

The EPA study (EPA 747/R-95/002b) discussed in Section 3.1.2 also evaluated chemical test kits. This study concluded that chemical test kits should not be used in lead paint testing as none of the test kits demonstrated sufficiently low rates of false positive as well as false negative classifications. However, such a requirement is probably not realistic for qualitative measures such as chemical test kits. The evaluated test kits with low false positive rates tended to have high false negative rates, and vice versa. The only way a chemical test kit can achieve low rates of both false positive and false negative classifications (e.g., less than 5 percent for each) is for its OC-curve to remain near zero for all lead amounts below 1.0 mg/cm^2, have a very steep slope at 1.0 mg/cm^2, and have a value near 1 for all lead amounts above 1.0 mg/cm^2. Since such performance is highly unlikely in practice, the way in which such methods are used might need to be reconsidered, as is suggested in the previous two sections.

Compared to an action level of 1.0 mg/cm^2 for paint, two test kits (LeadCheck and State Sodium Sulfide) had false negative rates of 6 percent and 1 percent, respectively. These rates are near the level required for 95 percent confidence using a negative screen. All other evaluated kits had higher false negative rates. Note that compared to an action level of 0.5 percent lead by weight, each test kit's false negative rate increased. However, the demonstrated overall false negative rates of 6 percent and 1 percent suggest that such chemical test kit technology might be appropriate as a negative screen. In conjunction with a decision tree, such a technology could be used to provide 95 percent confidence in the final statement made regarding a lead amount compared to an action level.

Compared to an action level of 1.0 mg/cm^2 for paint, one test kit (Lead Alert: Sanding) had a false positive rate of 9 percent, relatively close to a level required for 95 percent confidence using a positive screen. All other evaluated kits had higher false positive rates. Note that compared to an action level of 0.5 percent, the "Lead Alert: Sanding" test kit's false positive rate increased to 10 percent and the "Lead Alert: Coring" test kit's false positive rate was 11 percent. While some came close, none of the evaluated test kits appear to have demonstrated a false positive rate low enough to be used in a qualitative analysis decision tree for providing 95 percent confidence. For a detailed discussion and treatment of the chemical test kits mentioned in this section, refer to the field test report.

4.3.2 Field Investigation of On-Site Techniques (Chemical Test Kits)

In (Ashley, et. al., 1998-2), the performance of Rhodizonate-based chemical test kits for testing lead concentration in paint was examined. Results from on-site analysis using test kits were compared to a reference concentration determined AAS. OC curves of test kit response were given in the report, although they cannot be reproduced here as the data used in constructing them was not provided. The OC-curves show that these types of test kits are appropriate for negative screening; that is, they protect against negative results when the true lead concentration is greater than the action level. In particular, the curves indicate that, near 95 percent of the time, analyses using this technology typically will identify lead when the true lead level is at or above the action level.

Some information on false positive and false negative rates was provided in the report. Specifically, three false negative readings out of 66 samples with lead levels above 0.5 percent were recorded, for a false negative rate of 4.5 percent. Further, four false positive readings out of 105 samples where reference lead concentrations were below 0.06 percent were recorded, for a false positive rate of 3.8 percent. However, these rates may not reflect performance specific to lead levels very near the action level of concern. Instead, they reflect an overall performance averaged across concentrations either above the action level or below it. For example, many of the 105 samples with lead levels above 0.5 percent actually may have contained lead concentrations well above 0.5 percent, in which case the technology's performance is certain to be superior to what it would be near the action level. Thus, an overall false negative rate of 4.5 percent may not represent actual performance near the action level of 0.5 percent.

4.4 PRESENT AND DEFINE A DECISION TREE FOR 95 PERCENT CONFIDENCE, USING BOTH A NEGATIVE SCREEN AND A POSITIVE SCREEN

The discussion in Sections 4.1 and 4.2 above regarding qualitative measures used as negative and/or positive screens can be combined to form a model for making decisions with approximately 95 percent confidence. The use of such a decision tree would require access to both a negative screen qualitative measure and a positive screen qualitative measure. In absence of one or the other, the appropriate decision trees provided in the conclusions to Sections 4.1 and 4.2 instead could be employed.

The following two-way table diagrams the potential conclusions from an analysis for lead in paint, dust, or soil:

		Analysis Conclusion for Lead	
		Above Action Level	Below Action Level
True Lead Level	Above Action Level	Correct Conclusion	False Negative
	Below Action Level	False Positive	Correct Conclusion

A qualitative measure acting as a negative screen provides protection against false negative conclusions. A qualitative measure acting as a positive screen provides protection against false positive conclusions. However, even when armed with both a negative and positive screen, depending on the analysis results, a decision cannot always be made with 95 percent confidence. The following results can occur:

1. If both the negative and positive screen yield a negative indication for lead, then conclude with 95 percent confidence that the true lead level is below the action level.
2. If both the positive and negative screen yield a positive indication for lead, then conclude with 95 percent confidence that the true lead level is above the action level.
3. If the negative screen is positive AND the positive screen is negative, a decision regarding the true lead level as compared to the action level cannot be made with 95 percent confidence.
4. If the negative screen is negative AND the positive screen is positive, a rather conflicting result has been observed and the efficacy of one or both screens is in question. A decision regarding the true lead level as compared to the action level cannot be made with 95 percent confidence.

The fourth scenario should almost never occur in practice but is included for completeness. That is, a positive screen is much more likely to produce negative results; therefore, any time a negative screen yields a negative indication for lead, then almost certainly the positive screen will produce the same result. Similarly, a negative screen is much more likely to produce positive results; therefore, any time a positive screen yields a positive indication for lead, then almost certainly the negative screen will produce the same result.

The key to the limitation of qualitative measures in decision making is point number three above. Under this scenario, neither screen has provided sufficient evidence for a conclusion in one direction or the other regarding the action level. Therefore, such results would require some sort of definitive confirmation in order to make a decision with 95 percent confidence. However, this third scenario does provide some information. Such results suggest the presence of lead at a level that is probably not extremely far above the action level. In other words, with no lead present, the negative screen will yield a negative result 95 percent of the time. With a great deal of lead present, the positive screen will yield a positive result 95 percent of the time. Lead levels somewhere in between will tend to produce the result given by three.

From the above discussion, the decision tree in Figure 4.3 provides approximate 95 percent confidence that a correct decision is being made when using qualitative measures for lead analysis.

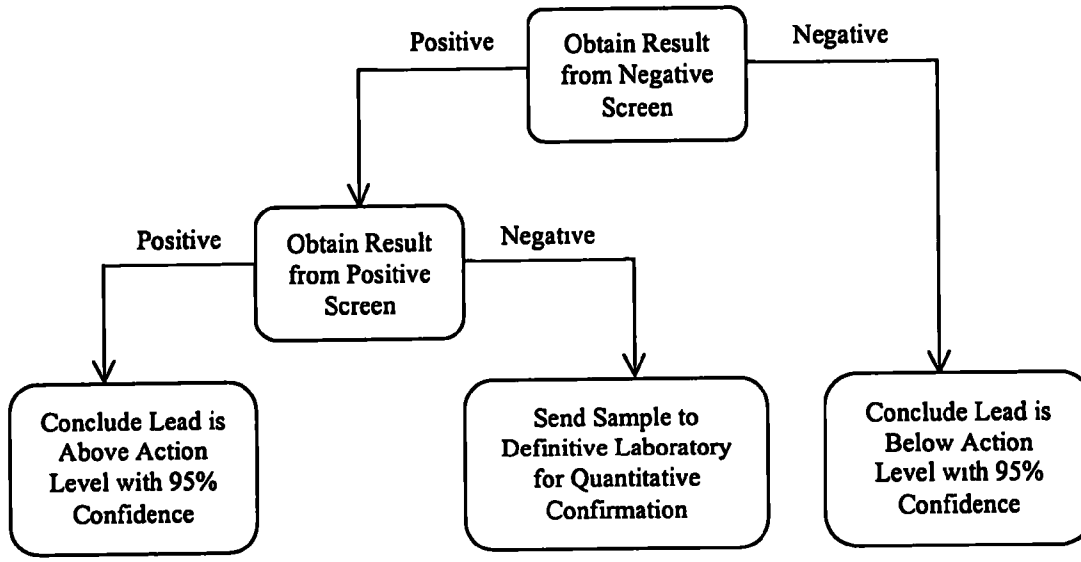


Figure 4.3 Decision tree for making statements with 95 percent confidence, using both a negative and positive screen.

Notice that the above decision tree is simply a combination of the decision trees given at the conclusions of Sections 4.1 and 4.2. The advantage of the above decision tree, assuming qualitative measures providing both negative and positive screens are available, is that it provides for the possibility of drawing a conclusion with 95 percent confidence in either direction of the action level, above or below. Of course, there is still the distinct possibility a conclusion with 95 percent confidence cannot be made; in which case more definitive information is required.

Finally, observe that the decision tree given in Figure 4.3 is not unique. First, the roles of the negative and positive screens could have been switched, producing a tree in which the positive screen is conducted first. Under this design, an initial positive result or an initial negative result followed by another negative result leads to a conclusion, and an initial negative result followed by a positive result requires more information for a decision to be made. Alternatively, the decision tree could have been designed symmetrically with both screens being performed at the initial step. In this case, two positive results or two negative results lead to a conclusion, while one negative result coupled with one positive result requires more information for a decision to be made. The advantage of the design in Figure 4.3 and the first one described in this paragraph is a reduction in cost due to implementation. Unlike these two asymmetric designs, a symmetric design that employs both screens at the initial step will always perform at least two analyses.

5.0 ADDITIONAL ISSUES AND CONCERNS

The purpose of this section is to identify and briefly discuss further issues that may require addressing during the process of expanding the NLLAP to cover the analytical technologies discussed in this report. The suggestions in this section are not meant to be rules. Instead, the goal of this section is to raise an important issue that needs to be addressed if NLLAP is to be expanded. Therefore, the provided suggestions are intended to initiate meaningful dialogue on this topic. Some issues arise due to the different types of entities that might be included in the scope of the NLLAP. Other issues come about due to the differing capabilities of various measurement technologies. But first, an alternative decision tree is offered, and its pros and cons are discussed.

5.1 AN ALTERNATIVE DECISION TREE FORMAT

One of the key aspects of the decision trees presented in the previous sections is the rule of defaulting to a definitive laboratory (testing firm) when an observed measurement lies within the gray-zone of a semi-quantitative or qualitative laboratory (testing firm). Such a rule treats definitive laboratories (testing firms) differently than their semi-quantitative and qualitative counterparts. The purpose of such a rule is to provide 95 percent confidence for the ultimate decision that will made, even in those cases when an initial result cannot do so. The idea is that the lack of precision associated with semi-quantitative and qualitative types of analyses will produce many instances in which an appropriate decision cannot be made with 95 percent confidence. Sending a sample to a definitive laboratory (testing firm) provides a safety net from which such a decision still might be made.

One possible alternative decision tree format would be to remove the rule of defaulting to a definitive laboratory (testing firm) when an observed measurement is not conclusively above or below an action level. With this approach, every laboratory or testing firm (definitive, semi-quantitative or qualitative) is viewed from the same perspective. That is, a definitive-type performer is not treated as a safety net from which semi-quantitative or qualitative laboratories (testing firms) can obtain a more reliable result.

Like the decision tree formats offered in the previous sections, under this alternative approach, each laboratory or testing firm has its own established performance characteristics (i.e., precision, bias and accuracy). However, the alternative format would indicate that a decision regarding an action level is made based solely on the results of the initial analysis – regardless of the type of laboratory or testing firm conducting the analysis. Thus, a result above a gray-zone is classified positive for lead with 95 percent confidence, a result below a gray-zone is classified negative for lead with 95 percent confidence, and a decision needs to be made when the result is in the gray-zone. One possibility for a gray-zone result would be to default to the conservative classification of positive for lead. This rule would be similar to the decision trees of this report for those cases when the initial result and the subsequent definitive result are both in gray-zones.

The advantage of this alternative decision tree format is its simplicity and reduction in time. Increased simplicity follows from the fact that the laboratory or testing firm performing the analysis does not have to determine when to send a sample to a definitive laboratory (testing firm). Furthermore, decisions can always be made based on initial results. The reduction in time is obvious, due to the time-savings associated with not having to wait for the definitive results.

This alternative decision tree format offers a reduction in laboratory analysis cost. However, such a cost reduction is likely trivial compared to the increased costs associated with unnecessary corrective measures that are dictated by the high rate of false positive results. That is, many times a sample that produces a gray-zone result in truth will contain a lead amount that is below the action level. Concluding the sample is positive for lead can lead to costly corrective measures that would not have been taken if a definitive analysis could have led to the proper conclusion.

Always defaulting to conservatively classifying gray-zone results as positive for lead will produce a higher rate of false positives. Similarly, always defaulting to classifying gray-zone results as negative for lead would produce a higher rate of false negatives which would be unacceptable from a public health and safety standpoint. Higher imprecision (e.g., semi-quantitative and qualitative analyses) leads to wider gray-zones, and subsequently more instances of some default decision needing to be made without an associated confidence level of 95 percent. Defaulting to a definitive laboratory in such instances increases the likelihood that the final decision is made with 95 percent confidence.

As a final note, a hybrid approach to the decision tree format presented in this report and the alternative format discussed in this section might be worth consideration as well. When gray-zone results are observed, this hybrid approach would default to obtaining additional definitive results unless the party of burden (e.g., the property owner) was willing to classify the result as positive for lead and take subsequent corrective action. This way, overall time and cost is reduced by not always requiring a further definitive analysis be done. Furthermore, the burden of increased false positives is incurred by choice. In summary, the property owner can either pay for the additional definitive analysis or conclude the gray-zone result is positive for lead and pay for any required corrective measures.

5.2 LABORATORY VERSUS TESTING FIRM DISTINCTION

The NLLAP Laboratory Quality Systems Requirements (LSQR) identifies minimum requirements for use by accreditation organizations to evaluate laboratories that perform quantitative analytical testing of paint chip, dust, and/or soil samples collected for lead analysis. These requirements include maintaining a data quality system, as well as guidelines on personnel, equipment, sampling methodology, and reporting. Any laboratory that meets the described requirements for lead analysis is described as accredited and is recognized for its ability to produce data of a recognized level of quality.

A testing firm is defined as an organization that may not meet the technical definition of a laboratory, but is still capable of producing data of a definitive-type quality. To become a definitive testing firm, the testing firm should be able to meet all of the requirements of the

NLLAP LSQR, but there may need to be a few possible exceptions. For example, one requirement of a laboratory is that a technical manager must be on staff, with a college degree in chemistry or a related science as well as three years non-academic laboratory experience. The technical manager function must be held by a laboratory employee and not contracted out. This requirement may be too costly in the case of smaller testing firms.

One possibility is that it may be more appropriate to have the technical manager position for testing firms be a part time position, where one technical manager works for several testing firms in a region. The job responsibility of the technical manager would not change, but this would allow the existence of smaller testing firms who perform straightforward lead screening analyses. That is, the burden of employing a full-time professional technical manager would be lifted. On a final note, another possibility would be to contract out the technical manager duties from an outside organization.

5.3 DIFFICULTIES IN ASSESSING PERFORMANCE

In some cases, for example portable XRF technology, laboratory control samples may not be available or relevant. One approach to this problem is to take split samples for screening type analyses. For example, five to ten percent of the analyses performed might require taking two samples from the same location, with one analysis done locally and the other sent to a definitive laboratory for verification. This is feasible for soil, but a potential problem arises for paint analyses. It is not always practical for paint chips to be removed from a residential unit and sent to a laboratory, as the owners of the unit may object due to the destructive nature of the sampling. Another approach is to include a standard reference material (SRM) with an on-site laboratory or testing firm, which can be done with portable XRF technology, and has the advantage of being non-destructive. However, it is not clear whether current SRM's applied to a technology like portable XRF can provide an appropriate characterization of precision and performance.

Finally, from a different perspective, characterizing the performance of a qualitative technology is not the same as that for a quantitative technology. For quantitative techniques, a standard error at an appropriate action level needs to be determined in order to calculate a gray-zone (i.e., 95 percent confidence interval). With qualitative techniques, an indication of positive or negative is provided. For such output, an appropriate tool for assessing performance is the operating characteristic (OC) curve. OC-curves give the probability of a positive indication, as a function of the true lead amount. For a detailed discussion on this issue and a methodology for deriving OC-curves, see (Koyak et al., 1998).

5.4 REPEATED SAMPLING TO IMPROVE PRECISION

Depending on the technology employed, on-site lead evaluation presents the opportunity to re-sample when faced with an initial gray-zone result. In this approach, more samples would be collected and analyzed when an initial result does not provide an answer with 95 percent confidence. For example, portable XRF technology offers the capability to obtain multiple measurements in a non-destructive fashion. Similarly, chemical test kits might offer this capability with only minimal impact on the surface being tested. However, if the sample error associated with, say, an XRF measurement is due mostly to the surface being tested, with little

error contributed by the instrument's measurement imprecision, then repeated measures will provide little benefit. That is, reducing the instrument measurement error through averaging repeated measures might not reduce potentially more significant components of error, such as error due to multiple layers of paint, error due to an uneven surface, etc. Given that this issue is beyond the scope of this report, performing repeated sampling under a semi-quantitative or qualitative analysis was not included in the decision trees provided in this report.

5.5 ALTERNATIVES TO THE GRAY-ZONES PROVIDED IN THIS REPORT

This section discusses two alternatives to dealing with gray-zones when making decisions based on analytical measurement. The first alternative abandons the concept of gray-zones entirely. The second uses a different method to calculate gray-zones.

5.5.1 An Alternative Approach that Avoids Gray-Zone Calculations

One alternative to using a gray-zone is to accept the instrument reading at face value. For example, if an instrument gray-zone is 0.8 to 1.2 mg/cm^2 and a recorded value of 0.9 mg/cm^2 is observed, then conclude that lead is not present above the action level, rather than sending the sample to a definitive laboratory for analysis. There are both advantages and disadvantages to this approach. The advantages are that this method saves time and resources. A decision can be reached quickly and cheaply. The disadvantage is that the probability of an incorrect decision being made will increase. The instrument gray-zone is designed to hold the probability of a false positive reading for unleaded components, or a false negative reading for leaded components, to less than or equal to 5 percent. If values inside the gray- zone are taken at face value, this protection against an incorrect decision is removed.

5.5.2 An Alternative Gray-Zone Calculation

A different method of calculating gray-zones is presented in (EPA 747-R-95-008). This method uses the concept of defining the boundaries of a gray-zone by deriving lower and upper threshold values (X_L, X_U). These threshold values are determined by a probability model, which assumes that lead levels are log-normally distributed. First, the probability that an XRF reading is below X_L given that lead levels are above the action level, divided by the probability of a true lead level being above the action level, is computed. This is the probability of a false negative reading. Then the probability of a false positive reading is computed by a similar method (an XRF reading being above X_U while the true lead level is below the action level). Finally, (X_U, X_L) are chosen to make these two probabilities both less than 5 percent.

Gray-zones calculated this way have the following features:

- The 5 percent targets for false positive and false negative classifications are an average across ranges of lead levels less than and greater than the action level, respectively. The probability of a false classification at a fixed lead level slightly above or below the action level is greater than 5 percent.
- The gray-zone is derived with the assumption that true lead levels have a log-normal distribution.

- The calculated gray-zones will be asymmetric, with the portion of the gray-zone above the action level typically being larger than the portion below the action level. This is a function of the log-normal distribution assumption.

The advantage of this method of calculating gray-zones is that it results in zones that are smaller than what is achieved by using the methods in this report. Hence, laboratories performing semi-quantitative analyses would have to rely on definitive laboratories for confirmation less often in practice. The disadvantages of this method are its complexity and the fact that it is not as conservative as the method presented in this report.

5.6 CONCLUSION

This report develops a field-decision performance based model for expanding NLLAP. This model takes the form of decision trees that provide 95 percent confidence in statements made regarding a lead amount relative to an action level. The provided decision trees vary according to the classification of laboratory or testing firm that is conducting the analysis. Ultimately, due to imprecision as represented by the concept of a gray-zone, exact 95 percent protection cannot be provided. However, in most cases, the decision trees provided in this report give near 95 percent protection against making incorrect decisions. For those cases when final results are inconclusive (i.e., in the gray-zone), it is recommended that a conservative conclusion be made, namely that the lead in the sample is above the action level. This approach protects those potentially at risk from a lead hazard. Finally, this report raises several issues critical to the expansion of NLLAP, which are yet to be resolved and require further careful consideration.

6.0 REFERENCES

(Ashley, et al., 1998-1) Ashley, K., Song, R., Esche, C., Schlecht, P., Baron, P., and Wise, T., "Ultrasonic Extraction and Portable Anodic Stripping Voltammetric Measurement of Lead in Paint, Dust Wipes, Soil, and Air: An Interlaboratory Evaluation" U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Cincinnati, Ohio, 1998.

(Ashley, et al., 1998-2) Ashley, K., Hunter, M., Tait, L., Dozier, J., Seaman, J., and Berry, P., "Field Investigation of On-Site Techniques for the Measurement of Lead in Paint Films" U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Cincinnati, Ohio, 1998.

(Ashley, 1995) Ashley, K., "Ultrasonic Extraction and Field-Portable Anodic Stripping Voltammetry of Lead from Environmental Samples", *Electroanalysis*, Vol. 7, No. 12, pp 1189-1190, 1995.

(Koyak, et al., 1998) Koyak, R., Schmehl, R., Cox, D., DeWalt, F., Haugen, M., Schwemberger, J., and Scalera, J., "Statistical Models for the Evaluation of Portable Lead Measurement Technologies - Part I: Chemical Test Kits", *Journal of Agricultural, Biological, and Environmental Statistics*, Vol. 3, No. 4, pp 451-465, 1998.

(White and Clapp, 1996) White, K., and Clapp, A., "Use of a field-portable anodic stripping voltammeter to determine the lead concentration of TCLP extracts", *American Environmental Laboratory*, October 1996.

(EPA 747/R-95/002b) U.S. EPA, "A Field Test of Lead-Based Paint Testing Technologies: Technical Report", EPA Report No. 747/R-95/002b, U.S. Environmental Protection Agency (EPA) Office of Pollution Prevention and Toxics, Washington, D.C. 20460, 1995.

(EPA 600/R-95/093) U.S. EPA, "Evaluation of the Performance of Reflectance and Electrochemical Technologies for the Measurement of Lead in Characterized Paints, Bulk Dusts, and Soils", EPA Report No. 600/R-95/093, U.S. Environmental Protection Agency (EPA) National Exposure Research Laboratory, Research Triangle Park, North Carolina, 1996.

(EPA 747/R-95/008) U.S. EPA, "Methodology for XRF Performance Characteristics Sheets", EPA Report No. 747/R-95/008, U.S. Environmental Protection Agency (EPA) Office of Pollution Prevention and Toxics, Washington, D.C., 1996.

(EPA LSQR Rev. 2.0) U.S. EPA, "National Lead Laboratory Accreditation Program: Laboratory Quality System Requirements (LSQR) Revision 2.0", U.S. Environmental Protection Agency (EPA) Office of Pollution Prevention and Toxics, Washington, D.C. 20460, 1996.

(EPA 600/R-97/145) U.S. EPA, "Environmental Technology Verification Report - Field Portable X-ray Fluorescence", EPA Report No. 600/R-97/145, U.S. Environmental Protection Agency (EPA) Office of Research and Development. Washington, D.C., 1998.

(ASTM E456-96) ASTM Standard Terminology E456-96, "Terminology Relating to Quality and Statistics."

(ASTM E1187-96a) ASTM Standard Terminology E1187-96a, "Terminology Relating to Laboratory Accreditation."

(ASTM E1605-95a) ASTM Standard Terminology E1605-95a, "Terminology Relating to Abatement of Hazards from Lead-Based Paint in Buildings and Related Structures."

APPENDIX A:

GLOSSARY

APPENDIX A:

GLOSSARY

There are numerous concepts that are referred to throughout this report. In order to understand the issues being presented, it is critical that such concepts are well defined and understood. As such, several key definitions are given below.

Accuracy: The closeness of agreement between a test result and an accepted reference value.

Action Level: A threshold lead level to which observed measurements are to be compared (e.g.. the federal threshold of 1.0 mg/cm^2 for lead in paint).

Analysis, Definitive: A quantitative analysis for the amount of lead in paint, dust, or soil with an associated accuracy satisfying the following constraint:

> The 95 percent coverage probability (confidence interval) of the analysis lies within ± 20 percent of the action level of concern.

Analysis, Semi-quantitative: A quantitative analysis for the amount of lead in paint, dust, or soil that does not satisfy the definitive analysis constraint.

Analysis, Qualitative: An analysis that does not provide quantitative information regarding the amount of lead in paint, dust, or soil, but does provide an indication regarding a presence or absence of lead above or below a specified concentration (e.g., a chemical test kit).

Bias: The difference between the expectation of the test result and an accepted reference value that is caused by systematic error.

Decision tree: A flow diagram providing guidance for making decisions, with a required level of confidence, regarding measured lead levels as compared to an action level.

Laboratory Accreditation: Formal recognition that a testing laboratory is competent to carry out specific tests or specific types of tests.

Laboratory or Testing Firm, Definitive: A laboratory or testing firm performing quantitative analyses that meets the defined definitive analysis criterion.

Laboratory or Testing Firm, Semi-quantitative: A laboratory or testing firm performing quantitative analyses that are not definitive in accuracy.

Laboratory or Testing Firm, Qualitative: A laboratory or testing firm performing qualitative analyses.

Note that laboratories might be of fixed-site, mobile facility, or field operation in nature (see NLLAP LSQR 2.0 for definitions). The distinction between laboratory and testing firm in this document is the requirement of a laboratory to employ a technical manager. No such requirement is made of testing firms.

Negative Screen: A qualitative analysis technique sensitive enough to provide an acceptable amount of protection (i.e., 95 percent confidence) against false negative errors.

Positive Screen: A qualitative analysis technique sensitive enough to provide an acceptable amount of protection (i.e., 95 percent confidence) against false positive errors.

Precision: Degree of mutual agreement between individual test results obtained under stipulated conditions.

Proficiency Testing: Determination of laboratory testing performance by means of interlaboratory test comparisons.

Repeatability: Precision under the following conditions: independent test results are obtained with the same method on identical test materials in the same laboratory by the same operator using the same equipment within short periods of time.

Reproducibility: Precision under the following conditions: test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment.

APPENDIX B:

DETAILS FOR DATA ANALYSES CONDUCTED

APPENDIX B: DETAILS FOR DATA ANALYSES CONDUCTED

The purpose of this appendix is to provide further detail regarding the statistical modeling and estimation of precision corresponding to different technologies of laboratories (testing firms). Included are details about statistical models that were used to determine estimates of precision at different action levels for paint, dust wipes, and soil.

B.1 ASSESSING THE PRECISION OF NLLAP RECOGNIZED LABORATORIES

ELPAT data for rounds 14 through 21 (N=32 means), as described in Section 2.1 of the main body of the report, were used to assess laboratory precision for NLLAP recognized laboratories. Means were treated as "true" values and the standard errors were assumed to be a linear function of the truth. Ordinary least squares regression was used to fit the following functional relationship:

$$\sigma = \beta_0 + \beta_1 * \mu \quad ,$$

where σ represents precision and μ represents the true lead level.

The results of analyses for paint, dust, and soil are given in Table B.1 below. Notice that the results for paint correspond to N=28 mean by standard deviation combinations, instead of N=32. Four observations with means well beyond the action level for paint of 1.0 mg/cm^2 were removed to more closely satisfy the assumed linear relationship between the ELPAT means and standard errors. Removal of these outliers was justified since these data did not represent NLLAP accredited laboratory precision near the action level. Furthermore, a more accurate portrayal of the linear relationship near the action level was obtained upon their removal. Satisfactory linear relationships were observed for dust and soil means with the removal of only one outlying data point in each data set.

Table B.1 Regression parameters for an NLLAP accredited laboratory's precision as a function of the true lead level.

Medium	Model: $\sigma = \beta_0 + \beta_1 * \mu$			
	β_0 (se)	95% CI for β_0	β_1 (se)	95% CI for β_1
Paint (n = 28)	0.002 (0.003)	[-0.001, 0.004]	0.063 (0.001)	[0.062, 0.064]
Dust (n = 31)	3.80 (1.218)	[1.413, 6.186]	0.066 (0.002)	[0.062, 0.070]
Soil (n = 31)	3.64 (1.180)	[1.322, 5.948]	0.052 (0.001)	[0.050, 0.054]

Descriptive statistics for the full and reduced data sets are provided in Table B.2. This table shows that the range of available data covers the action levels of concern, so use of the linear regression model to estimate standard deviations of measurement at the action levels is a valid approach.

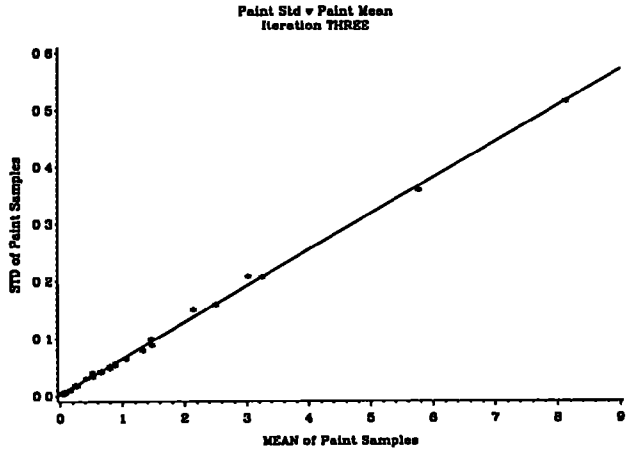
Table B.2 Descriptive statistics for paint, dust, and soil sample means for full and reduced data sets.

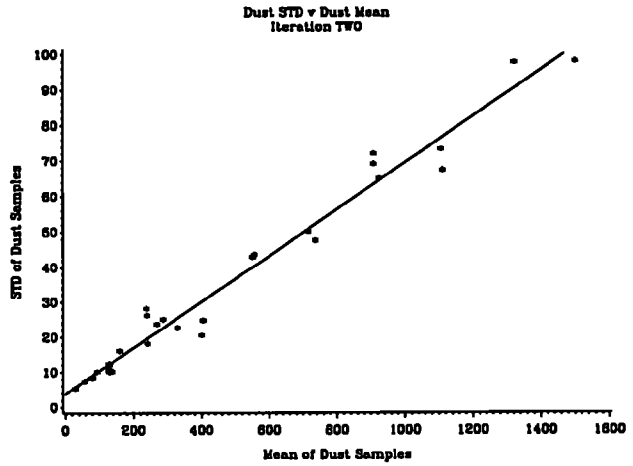
	N	Mean	Minimum	Maximum
Full Paint	32	1.813	0.0306	8.8
Reduced Paint	28	1.33	0.0306	8.1
Full Dust	32	465.6	29	1498.5
Reduced Dust	31	452.0	29	1498.5
Full Soil	32	889	34.8	3190
Reduced Soil	31	814	34.8	2788

The parameter estimates in Table B.1 can be used to calculate average gray-zone values representing NLLAP recognized laboratories. The model formula is used with the action level to calculate a precision value. Precision equals β_0 plus β_1 times the action level. The gray-zone is further calculated as the action level, plus or minus 1.96 times the precision. Table 2.1 of the main report presents a comparison between the definitive laboratory gray-zone requirement and the estimated gray-zone for NLLAP Laboratories based on the ELPAT rounds 14 through 21 data for NLLAP labs.

Based on Table 2.1, the definitive-type performance of the NLLAP recognized laboratories follows from the fact that the NLLAP recognized gray-zones are all narrower than the definitive gray-zone requirements, except for floor dust.

Figure B-1 below shows the ELPAT data corresponding to analyses of paint, dust, and soil. The vertical axis corresponds to the overall standard error associated with an analysis conducted by an NLLAP recognized laboratory. The horizontal axis contains the associated mean. The graphs shown indicate that an assumption of a linear relationship between laboratory standard errors and mean lead concentration indeed is reasonable.





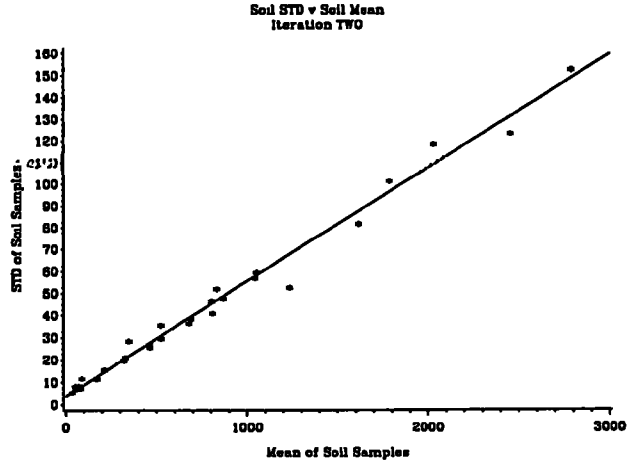


Figure B.1. Relationship between overall standard errors and means for NLLAP recognized laboratory analyses of paint, dust, and soil.

B.2 ASSESSING THE PRECISION OF PORTABLE XRF TECHNOLOGY

In 1993, a study was conducted by the U.S. Environmental Protection Agency (EPA) and the U.S. Department of Housing and Urban Development (HUD) to collect information necessary in order to establish federal guidelines on testing for lead in paint, (EPA 747/R-95/002b). (After the passage of Title X, Section 1017 of the Residential Lead-Based Paint Hazard Reduction Act of 1992, it became clear there was not enough information from existing studies to implement Title X). X-ray fluorescence (XRF) instruments were one of two field technologies analyzed.

In March and April of 1993 a pilot study was conducted in Louisville, Kentucky. From July through October of 1993, a full study was conducted in Denver, Colorado and Philadelphia, Pennsylvania. "A Field Test of Lead-based Paint Testing Technologies: Technical Report" was released in May 1995 and provides a complete technical report of this study.

The overall study objective was "to collect information about field measurement methodologies sufficient to allow EPA and HUD to establish guidance and protocols for lead hazard identification and evaluation." Included in "A Field Test of Lead-based Paint Testing Technologies: Technical Report" is the statistical model that was used to describe the relationship between XRF measurements and the lead level as analyzed by inductively coupled plasma-atomic emission spectrometry (ICP) (see Section 6.4.2). ICP is a method commonly used in laboratories to analyze lead in paint and is one of the techniques recommended for confirmation testing in the HUD Guidelines (United States Department of Housing and Urban Development (1990), "Lead-Based Paint: Interim Guidelines for Hazard Identification and Abatement in Public and Indian Housing," Office of Public and Indian Housing, Washington, D.C. 20410). This was one of many motivations for using ICP in the model (see Sections 3.3.1.1 and 6.4.2).

The statistical model is made up of two parts: a response component and a standard deviation (SD) component. The *response component* of the model mathematically describes the mean XRF reading at a particular level of lead. The linear function that is used to estimate the XRF response is described by the equation

$$a + b(Pb)$$

where:

- **Pb** represents the lead level in mg/cm^2 (as measured by ICP),
- **a** is the intercept and is compared with 0.0 to determine if the XRF readings are unbiased in the absence of lead, and
- **b** is the slope and is compared with 1.0 to determine if the instrument responded proportionately to changes in the lead level.

For example, a Lead Analyzer K-shell device was used to analyze paint chip samples from Brick. A linear model was fitted to the data with the parameters a = 0.084 and b = 0.703.

Therefore, when the lead level of a paint chip sample is analyzed as 0.82 mg/cm^2 by ICP, the mean lead level for the XRF instrument is predicted to be $0.084 + 0.703(0.82) = 0.66$ mg/cm^2.

The *SD component* of the model describes the variation in XRF readings at a particular level of lead. The non-linear function that is used to estimate the SD of the reading is described by the equation

$$[c + d(Pb)^2]^{1/2}$$

where:

- **Pb** represents the lead level in mg/cm^2 (as measured by ICP),
- **c** is the variance of XRF readings at a lead level of 0.0 mg/cm^2, and
- **d** is a measure of homogeneity of variance and is compared with 0.0 as the lead level increases to determine if variability remains the same as lead levels fluctuate.

For example, a model was fitted to the Lead Analyzer K-shell Brick data with parameters $c = 0.030$ and $d = 0.026$. Therefore when the lead level of a paint chip sample is analyzed as 0.82 mg/cm^2, the standard deviation for the XRF instrument is predicted to be $[0.030 + 0.026(Pb)^2]^{1/2}$.

Table B-3 contains a listing of the four model parameters **a**, **b**, **c**, and **d** for each of the six field portable devices tested in this study (Lead Analyzer, MAP-3, Microlead I, X-MET 880, XK-3, and XL). Since the Lead Analyzer and MAP-3 instruments could be operated by using either K-shell or L-shell X-rays, results were recorded twice for each of these instruments. Cells denoted with the symbol — indicate samples for which fitting a model was not deemed appropriate.

TABLE B.3. Estimated regression parameters for the response component and the SD component

DEVICE	SUBSTRATE	SAMPLE SIZE	MODEL PARAMETERS			
			a	b	c	d
Lead Analyzer K-shell	Brick (B)	93	0.084 (0.023)	0.703 (0 055)	0.030 (0.006)	0 026 (0.017)
	Concrete (C)	218	0 017 (0.012)	0.972 (0 054)	0 013 (0 002)	0 124 (0.032)
	Drywall (D)	111	-0.018 (0 009)	1.196 (0.115)	0.006 (0.001)	0.120 (0.081)
	Metal (M)	188	0.063 (0 021)	0.958 (0 055)	0.034 (0.006)	0.132 (0.035)
	Plaster (P)	218	0 030 (0.014)	0.861 (0 045)	0.019 (0.002)	0 037 (0 016)
	Wood (W)	351	0 013 (0.007)	1 266 (0.044)	0 007 (0 001)	0 180 (0.033)
Lead Analyzer L-shell	Brick (B)	92	0 038 (0.007)	0 036 (0.006)	0.001 (0 001)	0 0004 (0 0002)
	Concrete (C)	217	0 009 (0.001)	0 152 (0 012)	0 0001 (0 0002)	0.010 (0.002)
	Drywall (D)	113	-0 006 (0.001)	0 302 (0.029)	0 0000 (0.0000)	0.029 (0 008)
	Metal (M)	145	0.013 (0 002)	0.196 (0 023)	0.0002 (0 00006)	0.032 (0.007)
	Plaster (P)	213	0.002 (0.001)	0 201 (0 014)	0 0001 (0.0000)	0 019 (0 003)
	Wood (W)	337	-0.019 (0 001)	0.279 (0.016)	0.0002 (0.0000)	0.034 (0.005)
MAP-3 K-shell	Brick (B)	185	-0 599 (0 079)	0.797 (0.045)	0.857 (0.103)	0.012 (0.014)
	Concrete (C)	436	-0 661 (0.072)	1 212 (0 123)	0.807 (0.085)	0 182 (0.094)
	Drywall (D)	222	0.014 (0.040)	0.863 (0 209)	0.141 (0.018)	- 0 -
	Metal (M)	374	0 328 (0.039)	1.098 (0 071)	0 140 (0.019)	0.159 (0.049)
	Plaster (P)	443	-0.684 (0 065)	1.137 (0.102)	0 657 (0 069)	0 094 (0 067)
	Wood (W)	689	-0.052 (0.036)	1 410 (0.063)	0 239 (0 025)	0 203 (0 051)
MAP-3 L-shell	Brick (B)	183	0 012 (0.029)	0 109 (0 016)	0.055 (0.010)	0.003 (0 002)
	Concrete (C)	433	-0 141 (0 008)	0.201 (0.025)	0 008 (0 001)	0 019 (0 007)
	Drywall (D)	224	-0 115 (0 005)	0.498 (0.060)	0 002 (0 0003)	0 059 (0 021)
	Metal (M)	371	0 044 (0 037)	0.269 (0.055)	0.133 (0 018)	0.076 (0.024)
	Plaster (P)	426	-0.123 (0 010)	0.169 (0 029)	0 008 (0 001)	0.015 (0.007)
	Wood (W)	663	-0.079 (0.008)	0 425 (0 030)	0 008 (0.001)	0.068 (0.014)
Microlead I	Brick (B)	143	- -	- -	- -	- -
	Concrete (C)	218	283 (0 051)	1 094 (0 106)	0 375 (0 041)	0 143 (0 071)
	Drywall (D)	162	0.023 (0 031)	1.194 (0.175)	0.115 (0 013)	- 0 -
	Metal (M)	186	0 351 (0 060)	1.100 (0.075)	0 380 (0 053)	0 088 (0.050)
	Plaster (P)	415	0 010 (0 049)	1 068 (0 086)	0 265 (0.034)	0.118 (0 049)
	Wood (W)	348	0 001 (0.045)	1 424 (0 087)	0 389 (0 040)	0.448 (0 105)
X-MET 880	Brick (B)	72	- -	- -	- -	- -
	Concrete (C)	197	0 045 (0 003)	0.064 (0 013)	0.001 (0.0001)	0.004 (0.002)
	Drywall (D)	111	0.038 (0 002)	0 223 (0 031)	0.0002 (0.0003)	0.018 (0 006)
	Metal (M)	175	0.112 (0 017)	0.120 (0 032)	0.020 (0.003)	0.037 (0.008)
	Plaster (P)	210	0 048 (0 003)	0 072 (0.013)	0.0005 (0 0001)	0 006 (0.002)
	Wood (W)	334	0.042 (0.003)	0.259 (0.025)	0 0006 (0 0001)	0.083 (0.014)
XK-3	Brick (B)	143	0.861 (0.064)	1 016 (0 251)	0.359 (0.043)	- 0 -
	Concrete (C)	191	1.083 (0 063)	1.668 (0.227)	0 404 (0.043)	- 0 -
	Drywall (D)	112	-0.327 (0 040)	1 234 (0 254)	0 127 (0 019)	0.189 (0.363)
	Metal (M)	185	0.451 (0 058)	1.405 (0.140)	0.267 (0.037)	0.852 (0.192)
	Plaster (P)	215	0.535 (0.049)	1.035 (0.112)	0.298 (0.035)	0.103 (0 098)
	Wood (W)	342	-0.065 (0.035)	1 418 (0.073)	0.236 (0.024)	0.235 (0.064)
XL	Brick (B)	90	0.109 (0.016) *	0 183 (0.033)	0.016 (0 003)	0.018 (0.008)
	Concrete (C)	213	0 066 (0 009) *	0.391 (0 035)	0.008 (0 001)	0.051 (0.013)
	Drywall (D)	113	0.082 (0.019)	0.289 (0.109)	0 029 (0.004)	0 027 (0.035)
	Metal (M)	187	0.074 (0.017) *	0 546 (0.050)	0.020 (0.003)	0 132 (0 025)
	Plaster (P)	209	0 081 (0.011) *	0 405 (0.037)	0 010 (0.001)	0.017 (0.007)
	Wood (W)	191	0 049 (0.017) *	0.546 (0.037) **	0 008 (0.002)	0 091 (0.01)

* Nonparametric estimates reported. Standard error estimates obtained by bootstrapping.
** Estimates based on sample summary statistics for ICP < 0.1 mg/cm squared.

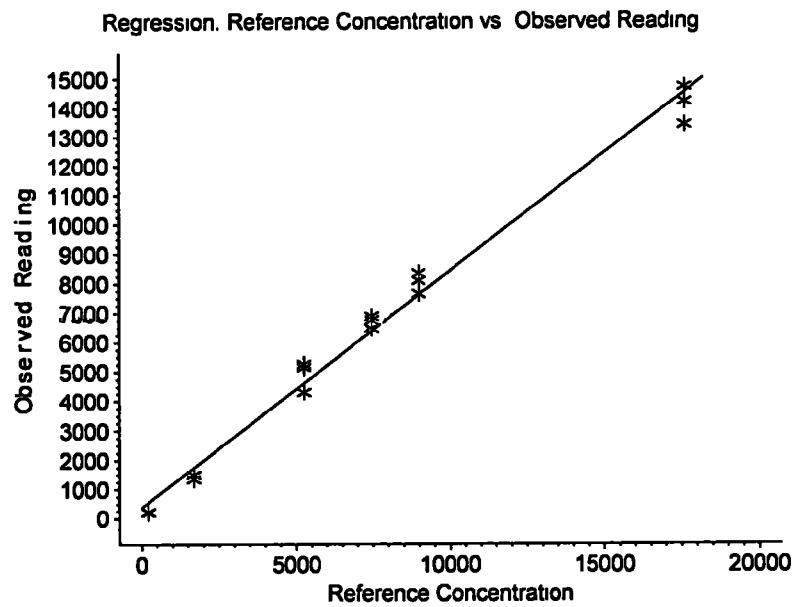
B.3 ASSESSING THE PRECISION OF UE/ASV TECHNOLOGY

The purpose of this section is to provide further detail regarding the statistical modeling of lead concentration readings taken by instruments using UE/ASV technology. Included are details about statistical models that were used to determine estimates of precision at different action levels for paint, dust and soil.

B.3.1 EPA Evaluation of the PaceScan 2000

As described in Section 3.2.1, the EPA evaluation analyzed UE/ASV performance at various concentrations of paint, bulk dust, and soil in (EPA 600/R-95/093) by taking repeated measurements of lead levels from a reference sample of known concentration (RTI Core Materials). For paint, 6 samples were analyzed using the PaceScan 2000 low range, and 7 samples were analyzed using the instruments high range (5 of these sample overlapped – used for both low and high ranges). Seven samples each were used for bulk dust and soil analysis.

Because the original data is presented in this report, it is possible to regress the true sample concentration onto measured concentration and get an accurate estimate of the variance of measurement at lead action levels. Lead concentrations are reported in units of μg/gram sample in this study, which can be converted to percent lead using the relationship 1 percent lead = 10,000 μg/g. A graph showing the relationship of measured concentration to true concentration for paint samples is shown below:

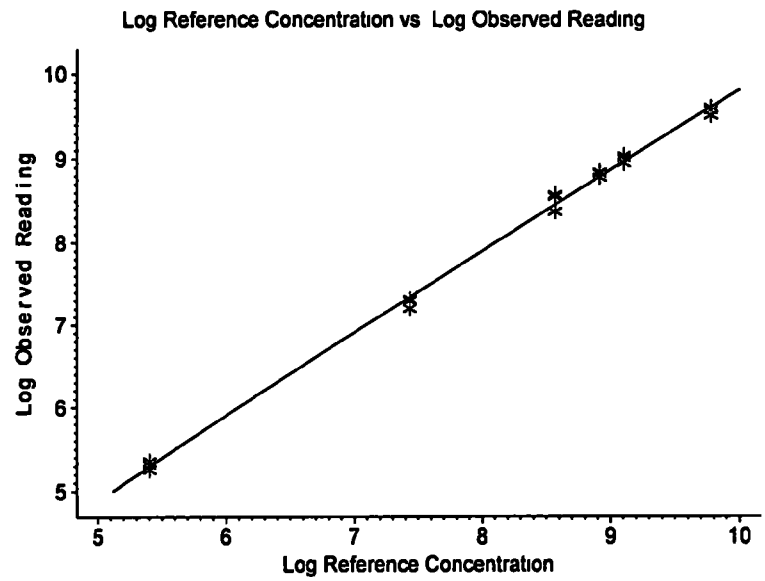


Source EPA 600/R-95/093 Paint Samples - PaceScan 2000. Low Setting

The action level for paint occurs at 0.5 percent lead, which is equivalent to 5,000 μg/g. Some problems arise when trying to fit a regression line to this data. The first is that variance is not uniform. The variance of observed sample readings is larger for high concentrations than for

low concentrations. The problem of non-constant variance was addressed by a log transformation of the data, regressing log true concentration onto log observed concentration. The second issue is that the data does not show a linear relationship. It was found that including a quadratic term in the model was appropriate.

A graph showing the relationship between log true concentration and log observed concentration is shown below:



Source EPA 600/R-95/093 Paint Samples - PaceScan 2000 Low Setting

This model provided a good fit for the data (R^2=0.9950) and was used for further analysis. Mean squared error is equal to 0.00450.

It is possible to use a transformation to estimate the standard error at the action level for paint of 5,000 μg/gram sample. This transformation is appropriate for converting results from log-normally distributed data to normally distributed data. Letting Y = observed concentration and T = true concentration, the above regression model assumes that:

$$\text{Log}(Y|T) \sim N(\mu, \sigma^2), \text{ where } \mu=\beta_0+\beta_1*\log T+\beta_2*(\log T)^2,$$
$$\text{and } \sigma^2 \text{ is estimated by MSE.}$$

If we wish to estimate the variance of Y|T instead of Log(Y|T), we can use the log-normal to normal transformation:

$$\text{Variance}(Y|T) = [\exp(\sigma^2)-1]*[\exp(2\mu + \sigma^2)]$$

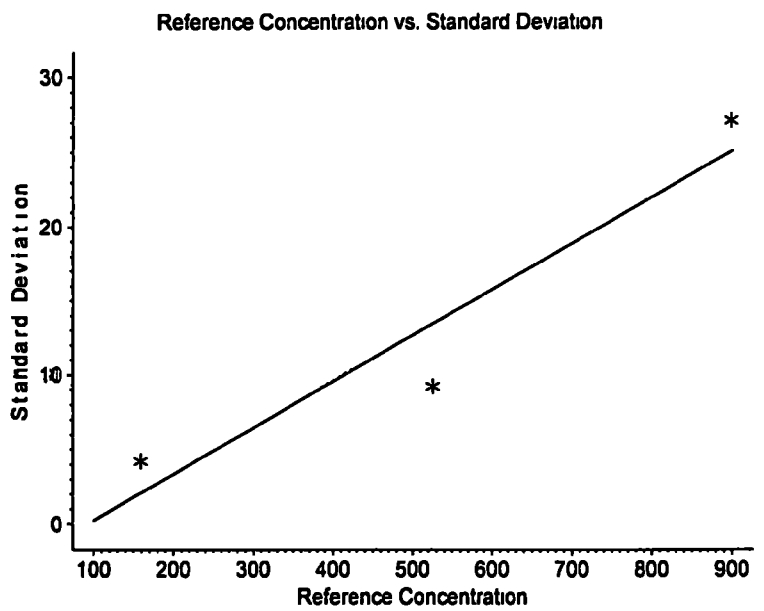
Applying this transformation gives an estimated standard error of 390 µg/g at the action level of 5,000 µg/g. Similar analyses were done for paint using the high setting, as well as for dust and soil.

B.3.2 Using Linear Extrapolation to Estimate Standard Deviations at the Action Levels of Concern

As described in Section 3.2.2, an interlaboratory evaluation studied reproducibility and reliability of UE/ASV measurements by comparing measurements made at 9 different laboratories (Ashley, et al., 1998-1). Three different samples of known lead concentration were provided to each lab, and two measurements were made on each sample per lab. A total of 18 measurements were taken on any given sample.

For paint, one of the samples provided had a lead concentration at the paint action level, so the reported standard error of the measurements for this sample was used.

For dust wipes, samples were not provided at the same concentration as the action levels. (The samples had concentrations of roughly 100, 200, and 900 µg/wipe). To get estimates for standard deviations of measurement at the applicable action levels, a linear projection using the known standard deviations at sampling concentrations was used. Standard error tends to increase with lead concentration in a linear fashion, illustrated in the graph below. Although only three data points are shown here, this linear relationship is typical of all studies considered.



Standard error at the 50, 250, and 800 µg/wipe action level was determined by using the regression line pictured above.

For soil samples, readings were taken at 500 and 3,000 µg/gram sample, whereas the action level of interest is at 2,000 µg/gram. Standard error of readings at the action level was

estimated using the same linear projection method as was used for dust wipes. A similar linear interpolation procedure was used to estimate standard deviations at action levels in section 3.2.4.

INTRODUCTION

- Programmatic need
  - Expand and redesign the current NLLAP.
- Study objectives
  - Evaluate lead analysis capability of laboratories/testing firms using portable XRF, UE/ASV and chemical test kits.
  - Construct decision models for different categories of laboratories/testing firms to be covered under an expanded NLLAP.

THE BOTTOM LINE

- The motivation in the approach to NLLAP expansion is to develop a system from which correct decisions (regarding environmental samples and measurements) are made 95% of the time.
- To this end, define a *gray-zone* as an action level of concern $\pm 2 \times$ laboratory precision.
- A *gray-zone* is to be interpreted as the area around an action level for which a decision cannot be made with 95% confidence.

THREE DISTINCT ANALYSES

- *Definitive*: A quantitative analysis with relative precision $\leq$ 10% at the action level of concern.
- *Semi-quantitative*: A quantitative analysis that does not meet the definitive requirement.
- *Qualitative*: An analysis that does not provide quantitative information, only a qualitative indication for lead (e.g., use of a chemical test kit).

SOME FINDINGS - QUANTITATIVE

- As expected, NLLAP-recognized laboratories, on average, perform at the *definitive* level (ELPAT rounds 14-21 data).
- Analyses employing UE/ASV technology demonstrate both *definitive* and *semi-quantitative* levels of performance (EPA evaluation of Pacescan 2000, NIOSH/Ashley studies).
- Analyses employing portable XRF technology most often perform at the *semi-quantitative* level (EPA field study, NIOSH/Ashley study, PCS's).

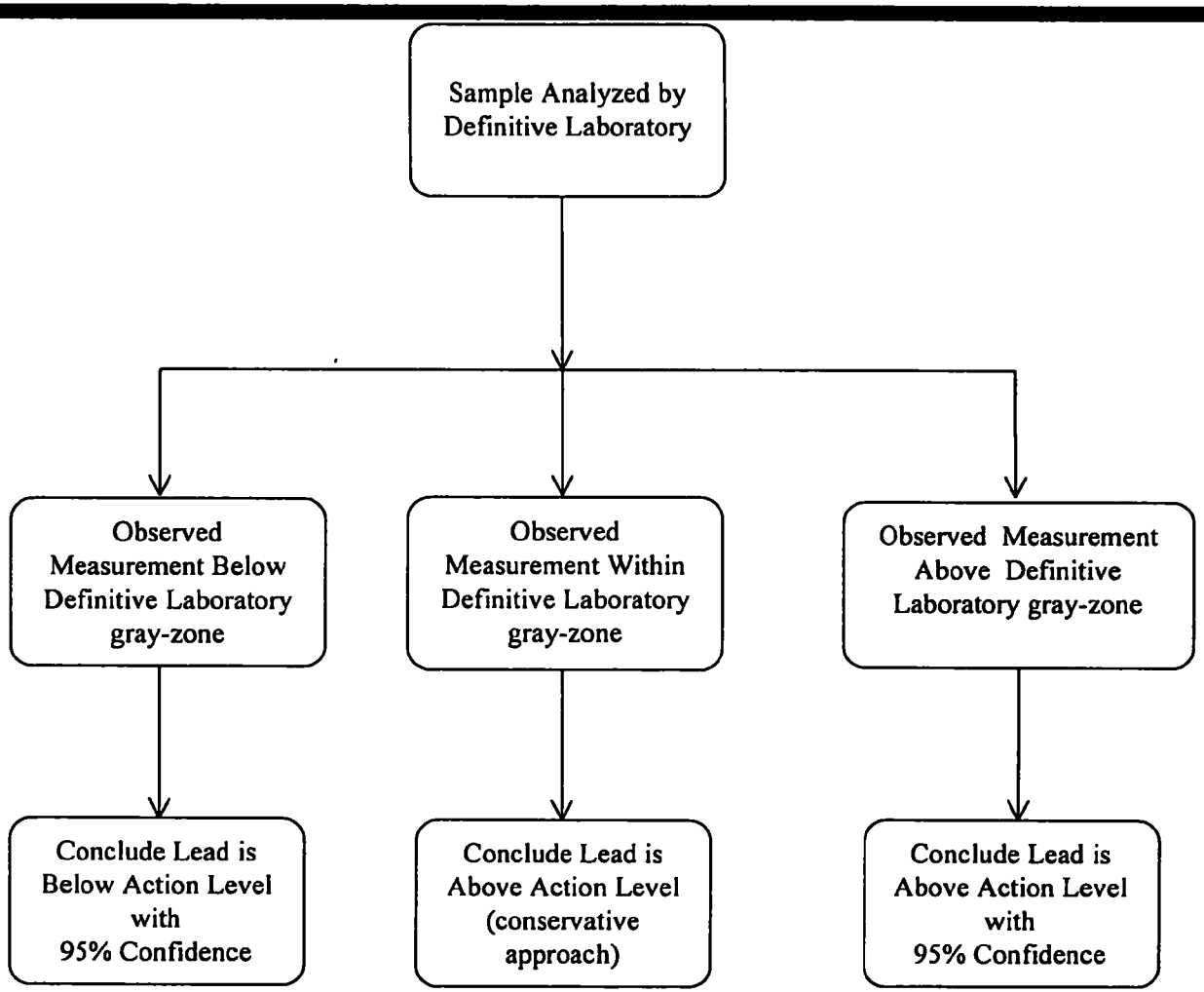
SOME FINDINGS - QUALITATIVE

- First, define two types of qualitative analysis
  - *Negative screen*: An analysis sensitive enough to provide 95% protection against false negative errors.
  - *Positive screen*: An analysis specific enough to provide 95% protection against false positive errors.
- Then, it was observed that (EPA field study)
  - Analyses using chemical test kits tended to demonstrate the sensitivity of a negative screen.
  - Analyses using chemical test kits typically lack the specificity necessary for use as a positive screen.

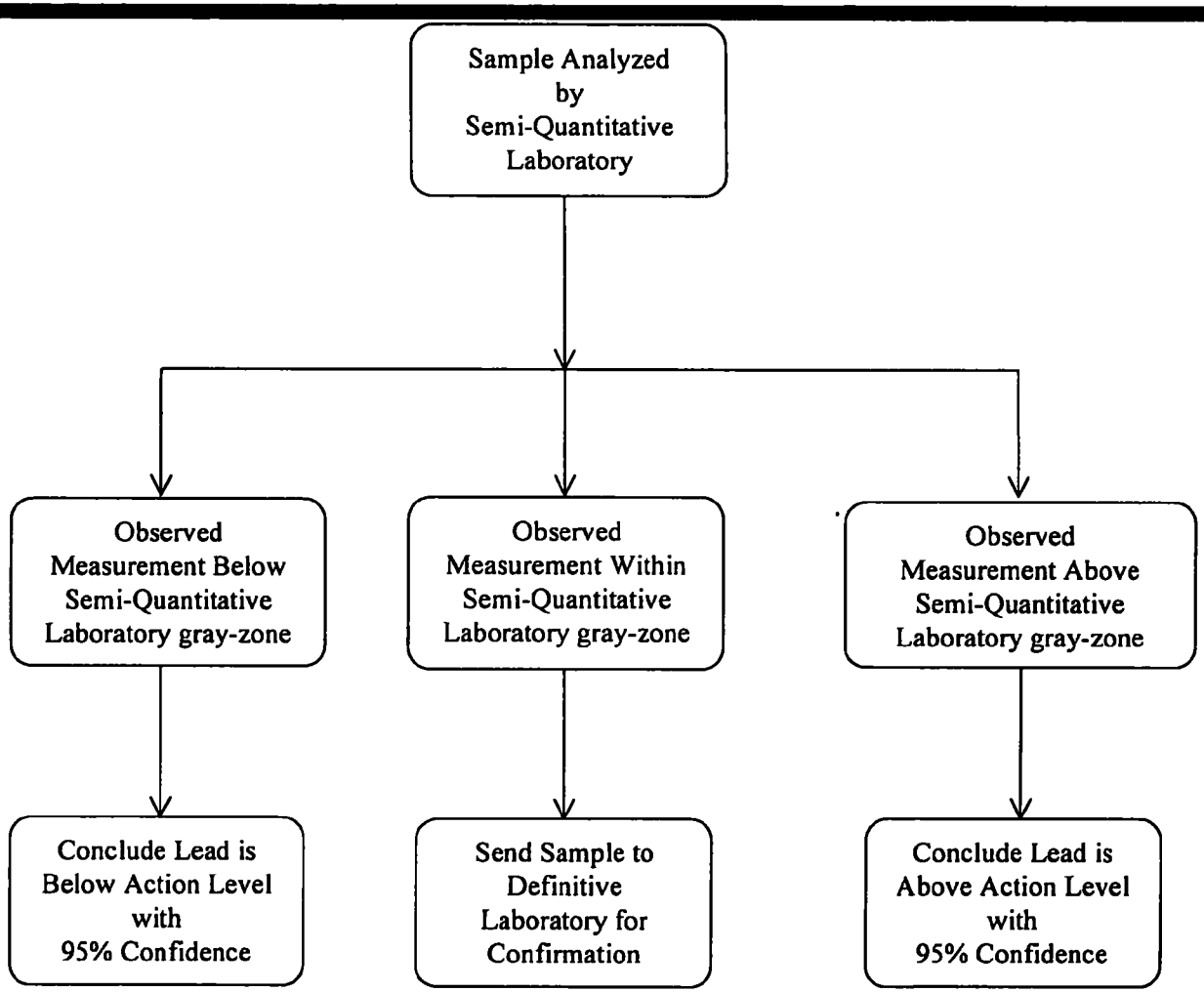
A MODEL FOR FIELD DECISIONS

- Consider decision trees for drawing conclusions in the field.
- In order to achieve the stated goal of 95% confidence in decision making, decision trees
  - are performance-based and therefore
  - depend on the type of analysis employed.

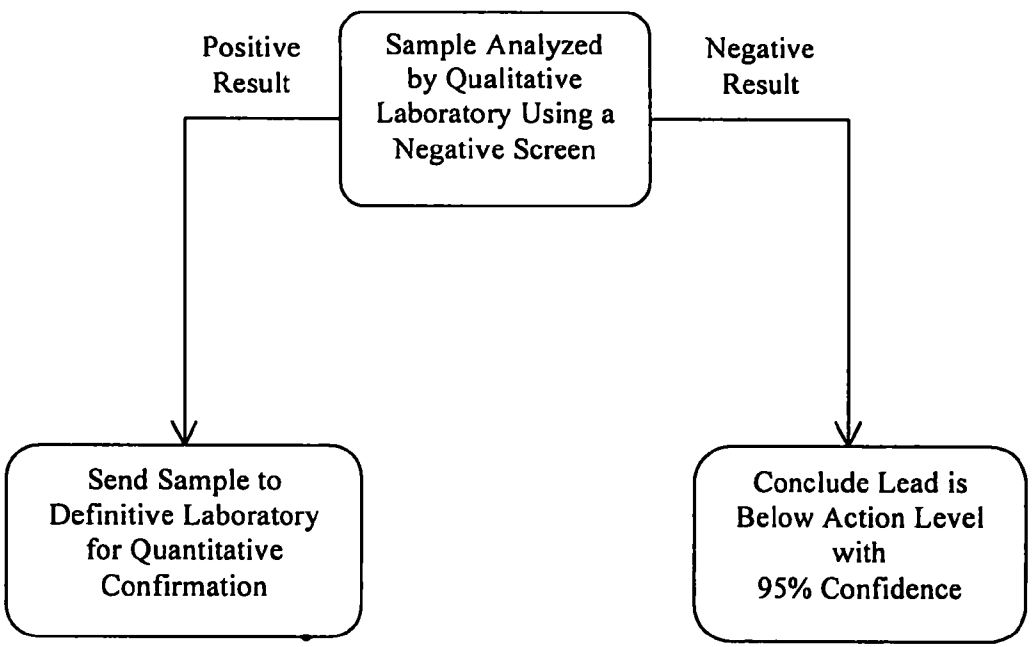
DEFINITIVE DECISION MAKING



SEMI-QUANTITATIVE DECISION MAKING



QUALITATIVE DECISION MAKING



SOME OPEN ISSUES

- Whether or not to default to a definitive laboratory for decision making
- Laboratory vs. testing firm distinction
- Difficulties in assessing performance (e.g., use of split samples or standard reference materials for XRF technology)
- Repeated sampling to improve precision (e.g., the average of multiple non-destructive XRF shots)
- Ignoring/avoiding gray-zone calculations