EPA 747-R-01-004
March, 2001

NATIONAL LEAD LABORATORY ACCREDITATION PROGRAM: DOUBLE-BLIND PROFICIENCY-TESTING PILOT STUDY

FINAL REPORT

Program Assessment and Outreach Branch
National Program Chemicals Division
Office of Pollution Prevention and Toxics
U.S. Environmental Protection Agency
401 M Street, S.W.
Washington, DC 20460

DISCLAIMER

FURTHER INFORMATION

Additional copies of this report can be obtained by calling the National Lead Information Center at 1-800-424-LEAD. Information about other technical reports on lead can be found on the Internet at the following address:
http://www.epa.gov/lead

This report is copied on recycled paper.

CONTRIBUTING ORGANIZATIONS

The study documented in this report was funded and managed by the U.S. Environmental Protection Agency (EPA). The study was conducted collaboratively by the following organizations:

Battelle

Under contract to the U.S. Environmental Protection Agency, Battelle was responsible for managing contractor efforts, establishing the study design, preparing and distributing the Quality Assurance Project Plan, recruiting study participants, establishing necessary standard operating procedures, preparing correspondence and instructions to study participants, reimbursing study participants for sample analysis costs, summarizing and analyzing the study data, and producing the final report.

American Industrial Hygiene Association (AIHA)

Under contract to Battelle, AIHA was responsible for providing consultation on accreditation programs, overseeing efforts to provide proficiency-test samples to study participants, acting as the "proficiency testing service" in this pilot (i.e., liaison with study participants), collecting the analytical results of the laboratory analyses, and providing necessary data to Battelle for data summary and analysis.

Research Triangle Institute (RTI)

Under contract to AIHA, RTI was responsible for preparing and storing proficiency-test samples for use in the pilot, receiving and storing sample containers from study participants, and distributing proficiency-test samples (with appropriate data tracking and reporting materials) to the study participants.

U.S. Environmental Protection Agency (EPA)

U.S. EPA was responsible for oversight in developing the study plan, managing and coordinating the study, performing recruitment of study participants, and reviewing and editing this report. The EPA Work Assignment Manager was John Scalera. Brion Cook was the Branch Chief of the Technical Branch under whose direction the study was conducted. The Project Officer was Sineta Wooten.

This page left blank intentionally.

**NATIONAL LEAD LABORATORY ACCREDITATION PROGRAM:
DOUBLE-BLIND PROFICIENCY-TESTING PILOT STUDY**

TABLE OF CONTENTS

Page

Table of Contents (cont.)

Page

List of Tables

Table Page

Table of Contents (cont.)

List of Tables (cont.)

Table | Page

Table of Contents (cont.)

List of Tables (cont.)

Table of Contents (cont.)

List of Tables (cont.)

List of Figures

Table of Contents (cont.)

List of Figures (cont.)

Table of Contents (cont.)

List of Figures (cont.)

This page left blank intentionally.

NATIONAL LEAD LABORATORY ACCREDITATION PROGRAM: DOUBLE-BLIND PROFICIENCY-TESTING PILOT STUDY

EXECUTIVE SUMMARY

This report documents the design, execution, results, and conclusions of a pilot study to gather information on how a double-blind proficiency-testing program could be incorporated within EPA's National Lead Laboratory Accreditation Program (NLLAP). Currently, the proficiency-testing program used within the NLLAP is the Environmental Lead Proficiency Analytical Testing (ELPAT) Program (NIOSH, 1994). The ELPAT Program is single-blind, in that the laboratories are aware that they have received a batch of proficiency-test samples for analysis, but they are unaware of the amounts of lead in these samples. In a double-blind program, the laboratory is unaware that its proficiency is being evaluated. Neither is the laboratory aware that proficiency-test samples are included within a batch of samples received for analysis. Therefore, a double-blind program is more likely than a single-blind program to characterize the overall performance of routine field sample analyses.

Twelve clients of NLLAP-recognized laboratories provided proficiency-test samples to laboratories in this double-blind proficiency testing pilot study and reported the results of analytical testing on these samples. These laboratories were found to be representative of NLLAP-recognized laboratories in their performance within Round 22 of the ELPAT Program. In each of three double-blind testing rounds, these clients were instructed to include the proficiency-test samples within their next available batch of field samples for shipment to one or more NLLAP-recognized laboratories for lead analysis. Within each double-blind testing round, 9 of the 12 clients who routinely collected dust wipe samples for lead analysis were provided with from 4 to 8 dust wipe proficiency-test samples to send to a particular laboratory. These samples were spiked with lead-dust, with half (i.e., from 2 to 4 samples) at a lower nominal lead amount than the other half. In addition, 10 of the 12 clients who routinely collected paint chip samples for lead analysis were provided with from 2 to 4 paint chip proficiency-test samples to send to a particular laboratory in a given double-blind testing round, where all paint samples contained the same nominal amount of lead.

The three types of dust-wipe proficiency-test samples analyzed in this double-blind pilot study (i.e., low-spiked dust wipes, mid-spiked dust wipes, paint chips) originated from three of the batches of proficiency-test samples that were prepared for Round 22 of the ELPAT Program. Therefore, the laboratories in this pilot study had previously analyzed one of each type of proficiency-test sample within Round 22 of the ELPAT Program. Generally, laboratories analyzed proficiency-test samples in the first testing round in this double-blind pilot study within 30 days of analyzing proficiency-test samples in Round 22 of the ELPAT Program.

Client recruitment involved submitting a news brief for publication in a monthly trade journal to the lead hazard control industry documenting the study and its objectives and inviting interested clients to contact the study team, and contacting risk assessors directly for their participation. The recruitment

process revealed a need to develop a strategy to educate laboratory clients on the benefits of a double-blind proficiency-test program, which would likely increase their participation in such a program.

Many clients encountered in the recruitment process were small organizations that either collected relatively few samples on a monthly basis for analysis, or had limited resources that made it difficult to participate in this study. For these reasons, and to improve the overall success rate of recruitment, clients were reimbursed for analysis costs associated with the proficiency-test samples they were provided in this study. In addition, materials such as sample containers and dust wipes were provided to some clients. The policy of providing such incentives to get laboratory clients to participate needs to be reviewed when establishing a double-blind program.

Although the three types of double-blind proficiency-test samples originated from the same batches of samples used in Round 22 of the ELPAT Program, the variability in the double-blind pilot study data is greater than the variability in the ELPAT Round 22 data for the same sample types and laboratories. In fact, while all of the laboratories that analyzed proficiency-test samples in the double-blind pilot study reported acceptable results within Round 22 of the ELPAT Program, 29 (approximately 11%) of the double-blind pilot study results were outside of the ranges that were considered acceptable within the ELPAT Round 22. Of these results, 10 were for low-spiked dust-wipe samples from four of the ten laboratories analyzing dust-wipe samples, 8 were for mid-spiked dust-wipe samples from four of the ten laboratories analyzing dust-wipe samples, and 11 were for paint chip samples from three of the ten laboratories analyzing paint chip samples. Thus, double-blind pilot study data for some laboratories may be more likely than single-blind study data to exceed the acceptance limits determined from data within the (single-blind) ELPAT Program.

Twelve of the double-blind pilot data values were labeled as statistical outliers (i.e., very high or very low values relative to other data of the same sample type in the given testing round). While some of these outliers were later determined to be the result of laboratories reporting invalid lead amounts, they were values that the laboratories reported to the clients and, therefore, would be used by the clients to make decisions on lead contamination. Therefore, statistical summaries and analyses were performed both with and without the outliers included. Other extreme data values that were reported inaccurately by the clients were revised upon obtaining laboratory reports and noting how the laboratory reported these values to the client.

Table ES-1 summarizes average lead amounts reported in the double-blind proficiency-test pilot study, by sample type and double-blind testing round. The target lead amounts associated with each sample type, as determined within Round 22 of the ELPAT Program, are the means specified in the last row of this table. Also included in Table ES-1 are ELPAT Round 22 data summaries for the group of laboratories participating in this pilot study, as well as for all 118 participating NLLAP-recognized laboratories (labeled as “reference laboratories” within ELPAT Round 22). Note that statistical outliers occasionally contribute to inflated standard

Table ES-1. Average (and Standard Deviation) of Measured Lead Amounts in the Double-Blind Pilot Study (by Testing Round) and in Round 22 of the ELPAT Program, by Sample Type

	Mean (Standard Deviation) (Sample Size)		
	Low-Spiked Dust Samples (μg lead)	Mid-Spiked Dust Samples (μg lead)	Paint Samples (% lead by weight)
Double-Blind Pilot Data			
DB Round 1	124.1 (39.4) (28)	265.6 (81.6) (28)	0.639 (0.124) (32)
DB Round 2	132.7 (29.8) (30)	276.2 (50.6) (30)	0.749 (0.340) (32)
DB Round 3	129.4 (24.5) (24)	282.1 (43.7) (24)	0.710 (0.204) (29)
Double-Blind Pilot Data, With Statistical Outliers Excluded (see Table 5-1)			
DB Round 1	133.4 (20.1) (26)	285.6 (36.8) (26)	0.658 (0.070) (31)
DB Round 2	128.9 (21.8) (29)	281.3 (43.1) (29)	0.659 (0.066) (29)
DB Round 3	129.4 (24.5) (24)	282.1 (43.7) (24)	0.656 (0.039) (27)
ELPAT Round 22 Data[1]			
Labs participating in the double-blind pilot	135.5 (10.1) (10)	284.2 (19.8) (10)	0.655 (0.056) (11)
NLLAP-recognized labs	127.9 (17.4) (118)	294.5 (270.2) (118)	0.726 (0.881) (118)
NLLAP-recognized labs (Winsorized data)[2]	129.0 (11.6) (118)	272.0 (24.4) (118)	0.6454 (0.040) (118)

[1] For a given sample type, one sample was tested per laboratory in Round 22 of the ELPAT Program.
[2] The specified means represented target lead levels for the specified sample types in Round 22 of the ELPAT Program.

deviations both in the double-blind testing rounds and in Round 22 of the ELPAT Program among NLLAP-recognized laboratories.

Even when statistical outliers were excluded, the standard deviation of the dust wipe proficiency-test sample data in each double-blind testing round was more than twice that reported in Round 22 of the ELPAT Program for the same group of laboratories (Table ES-1). While Levene's test indicated that, for each dust-wipe sample type, the differences in these standard deviations across the four testing rounds (the three double-blind testing rounds plus ELPAT Round 22) was not significantly different at the 0.05 level, the standard deviations of the log-transformed dust-wipe data (i.e., the data used in statistical analyses within this report) were significantly different at the 0.05 level across testing rounds. This was true even when statistical outliers were not excluded when calculating standard deviations within the double-blind testing rounds. The significant difference was primarily due to the lower variability observed in Round 22 of the ELPAT Program relative to the double-blind testing rounds.

Compared to the dust-wipe data, the double-blind proficiency-test paint chip sample data had standard deviations that more closely matched the standard deviation for paint chip sample data from ELPAT Round 22 (Table ES-1) when statistical outliers were excluded. These standard deviations did not differ significantly (at the 0.05 level) across testing rounds.

While it appears from Table ES-1 that the average double-blind dust-wipe proficiency-test sample results more closely matched the ELPAT Round 22 target levels in double-blind rounds 2 and 3 compared to double-blind round 1, the differences in average results for dust-wipes relative to their target levels were not significantly different across double-blind testing rounds at the 0.05 level for either sample type, based on tests performed within an analysis of variance. Furthermore, the analysis of variance concluded that when statistical outliers were excluded, the extent of variability in average laboratory results (i.e., lab-to-lab variability) did not differ significantly among the double-blind testing rounds, nor did the extent of variability in the results of multiple sample analyses within the same laboratory (i.e., within-lab variability) differ significantly among the double-blind testing rounds, at the 0.05 level. The same statistical conclusions were made on the results of paint chip sample analyses. This suggests that for each type of proficiency-test sample, with the exception of a few sample results that were labeled statistical outliers, the overall performance of the laboratories in this double-blind pilot study did not differ significantly across the three double-blind testing rounds.

Table ES-2 presents estimates of the percentages of total variability in the double-blind pilot study dat that was associated with lab-to-lab variation, according to double-blind testing round. This table shows that in testing rounds containing statistical outliers, lab-to-lab variability represented at least 90% of total variability. However, when statistical outliers were not present, lab-to-lab variability was generally about 70% of total variability for both types of dust-wipe proficiency-test samples, and slightly less than 50% of total variability for the paint chip proficiency-test samples. These latter percentages represent data over the entire study, as these percentages did not differ significantly (at the 0.05 level) across testing rounds when statistical outliers were excluded. However, these percentages were slightly higher when calculated for a specific double-blind testing round. Statistical acceptance criteria in a double-blind program designed similarly to this pilot study should consider both lab-to-lab and within-lab components of variation.

For each proficiency-test sample type, the overall average measurement reported by the laboratories within a double-blind testing round did not differ significantly across the three testing rounds (at the 0.05 level), and deviation of this average from the target level associated with the proficiency-test sample type was not statistically significant overall. This finding, along with the finding mentioned above that both lab-to-lab and within-lab components of variation did not differ significantly across testing rounds when statistical outliers were excluded from analysis, suggests that overall performance of the laboratories in this double-blind pilot study did

Table ES-2. Percentage of Total Variability in Double-Blind Pilot Study Results That Can Be Attributed to Lab–to-Lab Variation

Double-Blind Testing Round	Low-Spiked Dust Samples	Mid-Spiked Dust Samples	Paint Samples
All Double-Blind Pilot Data Included			
DB Round 1	99.6%	99.8%	55.8%
DB Round 2	64.7%	69.1%	94.0%
DB Round 3	85.2%	84.0%	98.8%
Statistical Outliers Excluded (see Table 5-1)			
DB Round 1	70.9%	80.5%	73.2%
DB Round 2	84.3%	80.8%	53.4%
DB Round 3	85.2%	84.0%	49.5%
All DB Rounds*	68.6%	68.7%	46.4%

* Provided as the percentages did not differ significantly (at the 0.05 level) across DB testing rounds.

not differ significantly across the three double-blind testing rounds for each proficiency-test sample type (when statistical outliers were disregarded).

Additional research is needed on identifying appropriate dust-wipe, paint, and soil materials for use in preparing proficiency-test samples for a double-blind program. The different types of wipes available for field dust collection can make it difficult to select a specific type of wipe for use in the proficiency-testing program. Paint-chip proficiency-test samples are finely ground, which makes them easily distinguishable from paint chip samples collected in the field, as field samples tend to be larger chips attached to pieces of substrate. Soil samples were not considered in this pilot study, because the small amounts of soil (1-gram) used to prepare proficiency-test samples are considered easily distinguishable from composite soil samples (5-10 grams) typically collected in the field.

This page left blank intentionally.

1.0 INTRODUCTION

1.1 BACKGROUND

Lead exposure has a well-known association with adverse health effects in humans, especially young children. To identify and control or abate lead hazards in the nation's housing, lead inspections and risk assessments are conducted under guidelines dictated by regulation under Section 402 of the Toxic Substances Control Act (TSCA) (40 CFR Part 745, August 29, 1996). In these activities, samples are collected of those environmental media that would most likely provide lead exposure to children when a lead source is present (e.g., dust, soil, paint chips). These samples are then transported to analytical laboratories to determine the amount of lead that is present in these samples.

In the early 1990s, the U.S. Environmental Protection Agency (EPA) began activities to develop a national laboratory accreditation program for laboratories involved in the analysis of lead in dust, soil, and paint chips. EPA's official recognition as the responsible federal agency for developing such a program was made within Section 405(b) of Title IV of TSCA. This program, known as the National Lead Laboratory Accreditation Program (NLLAP) (68 FR 38656; July 19, 1993), recognizes laboratories for their ability to analyze lead in dust, soil, and paint chips in support of efforts to identify, abate, and control lead-based paint and lead-based paint hazards.

In order to be recognized by EPA under the NLLAP, laboratories must achieve the following:

- Successful participation in the Environmental Lead Proficiency Analytical Testing (ELPAT) Program (NIOSH, 1994), a cooperative effort of the National Institute for Occupational Safety and Health (NIOSH) and the American Industrial Hygiene Association (AIHA).

- A successful systems audit of laboratory operations, conducted by a laboratory accreditation organization participating in the NLLAP.

Successful participation in the ELPAT Program is required for NLLAP-recognized laboratories. The ELPAT Program is a *single-blind* proficiency testing program, where the proficiency testing service sends proficiency-test samples directly to the laboratories for analysis. Therefore, while the laboratories are unaware of the amount of lead in samples that are part of the program's performance evaluation, they are aware of when they are analyzing such samples.

1.2 PURPOSE FOR WORK

The need to perform lead analyses within the rapid-response, high-volume, cost-cutting environment common in many laboratories can negatively impact a laboratory's routine performance. Therefore, laboratories may try to minimize this impact when analyzing a batch of proficiency-test samples. This action is possible in a *single-blind* proficiency-testing program, such as the current

ELPAT Program, in which the laboratories are aware of when they are analyzing proficiency-test samples.

By contrast, a *double-blind* proficiency testing program would incorporate proficiency-test samples blindly within batches of field samples provided by clients of the laboratory. Because the proficiency-test samples would not be identified as such in these batches, laboratories would ideally not recognize such samples as proficiency-test samples and would therefore treat these proficiency-test samples with the same degree of care and attention they use in handling and analyzing routine field samples. Thus, variability associated with field sample testing may be better estimated by the results of double-blind testing rather than single-blind testing. In addition, a double-blind program can increase the level of alertness by the laboratories concerning quality, as they would not know when double-blind proficiency-test samples may arrive for analysis.

In determining whether to consider a double-blind proficiency testing program as a supplement to the ELPAT Program within the NLLAP, it was necessary to conduct a pilot version of a double-blind program to evaluate logistical considerations, to develop appropriate protocols, and to establish statistical performance criteria.

1.3 OBJECTIVES

The objectives of this pilot study were to:

- Design and test a protocol for double-blind proficiency testing under the NLLAP
- Compare analytical performance of NLLAP-recognized laboratories between the double-blind pilot study and the (single-blind) ELPAT Program.

Note that this report does not suggest appropriate acceptance criteria for double-blind proficiency-test samples, but instead provides important information to be used by those who must determine such criteria.

1.4 REPORT ORGANIZATION

Chapter 2 documents the study design used in the pilot and how it was implemented. Procedures used to recruit study participants, the outcome of the recruitment process, and experiences encountered in recruiting and working with the study participants and in obtaining the study data are presented in Chapter 3. The information in Chapter 3 provides valuable information in determining how to organize and operate a double-blind proficiency test study. Chapter 4 presents the methods used to perform data analyses on the study data, and results of executing these methods on the pilot study data are presented in Chapter 5. Quality assurance issues, such as sample fabrication, data management and sample tracking, and data quality checking are discussed in Chapter 6. Finally, conclusions and recommendations made from executing this pilot study and in support of the study objectives are presented in Chapter 7.

Research Triangle Institute (RTI) prepared the proficiency-test samples and performed sample transfer and shipment to the study participants. A report on RTI's responsibilities on this pilot study is found in Appendix A.

This page left blank intentionally.

2.0 STUDY DESIGN AND OPERATION

A double-blind proficiency-testing program makes every effort to conceal from the laboratory the identity of proficiency-test (PT) samples, presenting them as field samples. Therefore, it was necessary to design this pilot study and to establish procedures to minimize the likelihood of laboratory recognition of its participation in this study. This chapter presents the study's underlying design to support the objectives presented in Chapter 1. This chapter also presents how procedures such as sample distribution and analysis reporting were implemented.

The overall organization of the double-blind pilot study, including the organizations involved in implementing the study, is illustrated in Figure 2-1. This figure indicates that the "proficiency-testing service" (PTS) of this pilot study, AIHA with support from RTI, was the primary point of contact with the study participants ("Clients"), who in turn were the points of contact with the laboratories who analyzed the proficiency-test samples. As a means of comparison, this figure also contains the flow of materials and information that occurs in each round of the single-blind ELPAT Program, which involves the above participants as well as NIOSH.

To allow for comparisons to single-blind results and to alleviate any sample integrity concerns among laboratories, operation of this double-blind pilot study was coordinated with Round 22 of the ELPAT Program. The proficiency-test samples analyzed within the pilot study consisted of the same batches of materials used to prepare the proficiency-test samples for Round 22 of the ELPAT Program. The proficiency-test samples analyzed within the pilot study consisted of the same bulk test material that was used in preparing the proficiency-test samples for Round 22 of the ELPAT Program. Furthermore, the PaceWipe™ brand of dust-wipe, used in the ELPAT Program, was also used in this pilot study. The PaceWipe™ is similar in composition to a pre-packaged moist towelette provided by restaurants for hand cleaning. This would be likely to reduce concerns from laboratories that the double-blind samples were prepared differently from single-blind samples, which could have led to more highly variable results.

2.1 STUDY PARTICIPANTS AND TESTING ROUNDS

As discussed in Chapter 3, this pilot study recruited lead-based paint inspectors and risk assessors who contract with NLLAP-recognized laboratories to perform lead analysis on the environmental samples (dust and paint) that they collect from residences. The initial goal of the study design was to enroll enough laboratory clients so that from 10-15 different laboratories would each have two clients submit a given type of proficiency-test sample for analysis for each test round. This would permit lab-to-lab variation in the analytical results to be estimated and would provide information on how performance at a given laboratory may differ from client to client. It would also reflect a design option for a double-blind program that would reduce a laboratory's vulnerability to having its performance negatively impacted by a specific client's actions. However, while each type of proficiency-test sample was analyzed by at least 10 different laboratories in this study, most laboratories received samples from only one client.

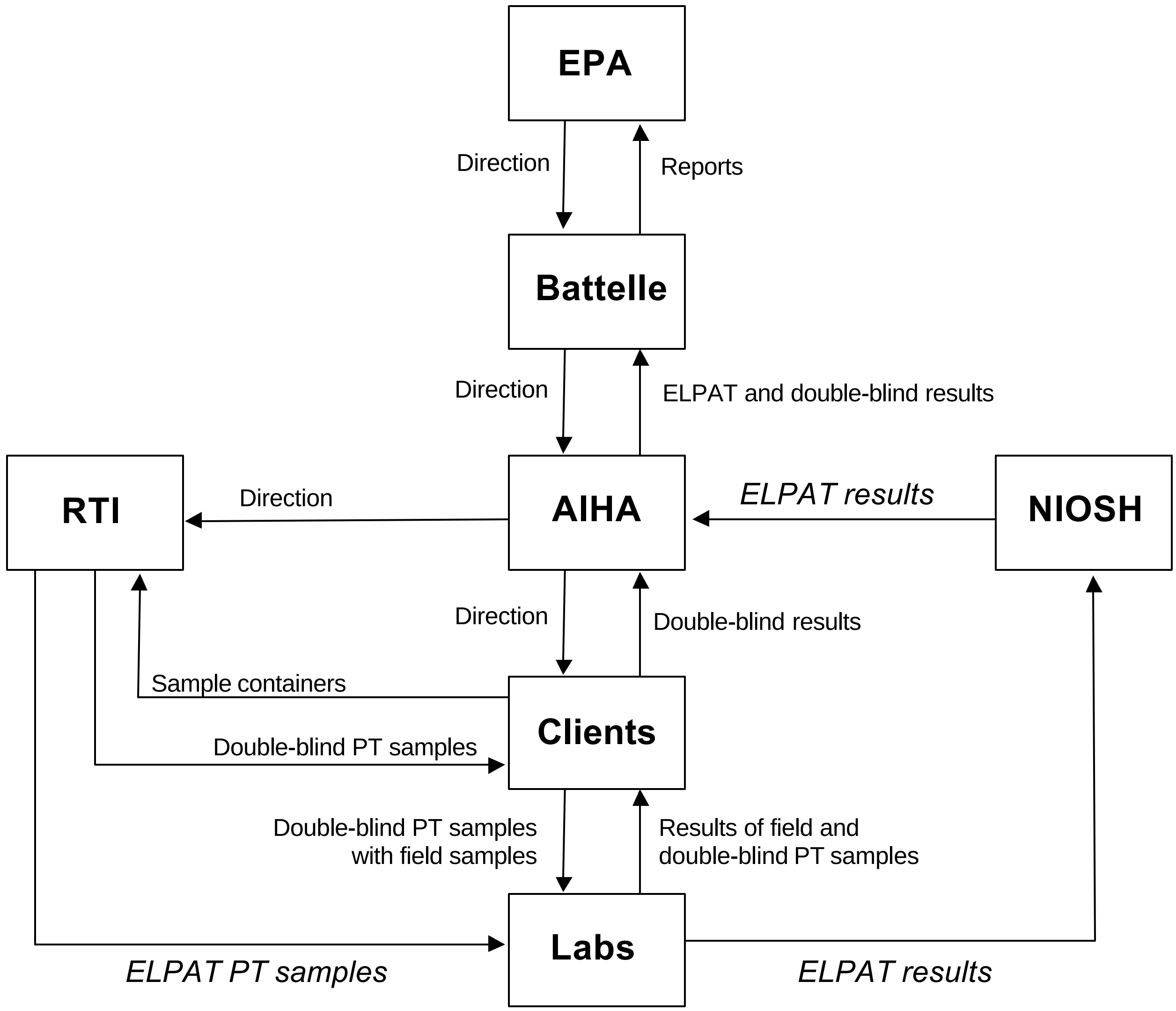

Figure 2-1. Organization of the Double-Blind Proficiency Testing Pilot Study, and Relationships Between Participating Organizations

This pilot study consisted of three testing rounds. The first testing round was initiated within one month of laboratories analyzing samples in Round 22 of the ELPAT Program, and the second and third rounds occurred approximately one and three months, respectively, following the first round. Proficiency-test samples analyzed in the second and third rounds were analogous to the samples submitted in the first round. Within each round, each recruited client was supplied with proficiency-test samples, with instructions to place these samples into their next available batch of regular field samples (of the given sample type) and to submit them to the laboratory for analysis. Incorporating multiple testing rounds in this study provided information on how laboratory performance in a double-blind proficiency-testing program may vary over time and over different analysis conditions (e.g., different technicians, calibrations). In particular, it was of interest to determine whether laboratory performance improved over time, especially if a given laboratory eventually suspected that they were being "double-blinded."

2.2 TYPES OF PROFICIENCY-TEST SAMPLES

Two different types of proficiency-test samples were considered in this pilot study: **dust-wipes and paint chips**. The wipe method for dust collection was considered for two reasons:

- The wipe method is most often used by risk assessors to collect dust samples within a residence.

- The rule (in response to Section 403 of Title IV of TSCA) that establishes criteria for determining when household dust is considered lead-contaminated (40 CFR Part 745; January 5, 2001) assumes wipe techniques for sampling household dust for lead analysis.

Paint chip samples were considered for three reasons:

- They are usually collected for laboratory analysis whenever *in situ* methods (e.g., x-ray fluorescence, or XRF) provide inconclusive results on the presence of lead in paint, or when such methods cannot be used (e.g., when paint is found on certain curved or ornate surfaces).

- Risk assessors may collect paint chip samples from deteriorated paint surfaces while doing a risk assessment.

- To evaluate the feasibility of including paint chip samples within a double-blind testing program.

Within Round 22 of the ELPAT Program, proficiency-test samples were prepared (using methods described in Section 6.1) at four different lead levels within each of three matrices (dust, soil, paint chip), thereby representing 3x4=12 sample types. Proficiency-test samples prepared within three of these sample types were selected for use in the double-blind pilot program. These sample types are documented in Table 2-1 and consist of dust-wipe samples at the two lowest lead levels and paint-chip samples at the second-lowest lead level. The lower two dust-lead levels were selected because the ELPAT Program typically observes an increased failure rate at low lead levels (Schlecht et al., 1996), and the Federal action level at the time of ELPAT Round

Table 2-1. Information on the Three Types of Dust and Paint Proficiency-Test Samples Prepared in Round 22 of the ELPAT Program and Used in the Double-Blind Pilot Study

Sample Type ID Within ELPAT Round 22	Matrix/Lead Level	Target Lead Amount[1]	Acceptance Range for Individual PT Sample Results Within ELPAT Round 22[2]
22W2	Dust-wipe/lower level	129 µg	94 - 164 µg
22W3	Dust-wipe/higher level	272 µg	199 - 345 µg
22P4	Paint-chip	0.6454%	0.5264 - 0.7645 %

[1] Mean of Winsorized data for 118 NLLAP-recognized ("reference") laboratories within ELPAT Round 22 (as documented in the round's individual laboratory reports), with each laboratory analyzing one sample of the given matrix and lead level.
[2] Plus and minus three standard deviations of the target lead amount, as calculated from Winsorized data for the 118 reference laboratories within ELPAT Round 22 and as documented in the round's individual laboratory reports.

22 was 100 µg/ft^2. The selected paint-lead level is within the range at which some *in situ* XRF instruments yield inconclusive results for the presence of lead (USEPA, 1995; USHUD, 1995).

Table 2-1 also includes the following for each sample type:

- the "target lead amount," or the lead amount to which each result of the given sample type was compared in this study's statistical analysis

- the "acceptance range," within which ELPAT Round 22 proficiency-test sample results were categorized as "acceptable" for the given sample type.

These two columns were calculated within Round 22 of the ELPAT Program using Winsorized data for the118 laboratories in this testing round that were NLLAP-recognized (labeled as "reference laboratories)."[1] The target lead amount is the mean of these data, and the acceptance range is plus and minus three standard deviations of this mean.

Paint chip samples used in the ELPAT Program can be difficult to use as double-blind proficiency-test samples because they are ground to a small particle size (<120 µm) during the material homogenization stage of sample development. Paint chips collected in the field are typically coarser and more likely to contain larger pieces of the substrate compared to proficiency-test samples. Therefore, to reduce laboratory suspicion of the paint-chip proficiency-test samples, clients within this

[1] Reference laboratories are the NLLAP-recognized laboratories going into the given testing round and are accredited by AIHA. Winsorization involves replacing the highest 5% of the 118 data points (for a given sample type) with the maximum of the remaining data, and the lowest 5% of these data with the minimum of the remaining data.

double-blind pilot study were instructed to refer to these samples as either "paint chips removed from a brick- or concrete-painted surface" or as "dust contaminated by paint."

While the ELPAT Program also includes proficiency-test soil samples, and soil is collected by many risk assessors, soil samples were not included among the proficiency-test samples in the double-blind pilot study. This is because proficiency-test soil samples in the ELPAT Program are typically less than one-half of the mass of soil samples collected in the field and, therefore, could be readily identified as proficiency-test samples. As field soil samples often exceed 10 grams, proficiency-test soil samples of mass similar to typical field samples would be prohibitively expensive to prepare.

2.3 NUMBERS OF PROFICIENCY-TEST SAMPLES PROVIDED TO EACH CLIENT

Within each testing round, each client received at least two proficiency-test samples of a given type (dust wipe, paint chip). Clients were provided with one or both types of samples, depending on what types of samples they typically provide to NLLAP-recognized laboratories monthly. By having at least two analytical results for each type of proficiency-test sample for a given client in a given testing round, within-laboratory variability in the analytical results could be estimated within each testing round.

In some situations where a laboratory received a given type of proficiency-test sample from only one client, that client was provided with four samples of the given type, rather than two. The additional samples provided additional information on within-laboratory variability and permitted each laboratory to analyze the same total number of samples in a given round, and therefore, to provide the same amount of information.

Specific numbers of proficiency-test samples provided to each participating client (and forwarded to each laboratory) are provided in Section 3.3.

2.4 SAMPLE TRANSFER TO CLIENTS AND LABORATORIES

The flow diagram in Figure 2-1 illustrates how the proficiency-test samples in this double-blind pilot study were transferred from representatives of the proficiency-testing service (RTI) to the laboratory clients, who then sent the samples to the laboratory for analysis. The study design had to include procedures for the various transfers of these samples from one participant to another, while retaining their proper identification, limiting their ability to be identified as proficiency-test samples, and not compromising their lead content. These procedures are discussed briefly below.

2.4.1 Transferring Samples to Client-Supplied Sample Containers

Placing proficiency-test samples in the same type of sample container that a client uses for regular field samples was necessary to ensure that the proficiency-test samples remained blind within a batch of regular field samples. As a result, the clients were requested to provide the proficiency-testing service with a specified number of uncontaminated sample containers that they normally used when submitting samples to an NLLAP-recognized laboratory. Clients were notified of this requirement during the recruitment process. The number of containers to supply was included in the letter the clients received when they were selected for the study (Appendix C). The types of sample containers received from the clients included glass jars, plastic jars, centrifuge tubes, and plastic bags.

Upon verifying that sample containers from participating clients were free of lead contamination (Section 6.1.3), the proficiency-test samples (paint chips and PaceWipes™ spiked with leaded dust) were transferred from the plastic scintillation vials in which they were initially stored after preparation to the client-supplied sample containers. An analytical method was applied to verify the quantitative transfer of samples from one container to another (Section 6.1.4).

2.4.2 Shipping Samples to Clients

After placing each proficiency-test sample into client-supplied sample containers, each sample was assigned a unique identification number (the "PTS Sample ID"). The form of the PTS Sample ID is provided in Section 5.1 of Appendix A. This ID number was printed on a label which was placed onto a clean, plastic bag. Each sample was then placed into its plastic bag, and the bag was sealed. At this point, the double-blind proficiency-test samples were ready for shipment to the clients. All samples shipped in the three rounds of the study were prepared and identified prior to the first testing round.

The proficiency-testing service shipped the proficiency-test samples to the recruited clients via priority mail in three distinct testing rounds. This approach differed from the initially-proposed approach which was to provide samples for all three testing rounds in one shipment at the beginning of the study. In the original plan, the clients were to store the samples appropriately until they were notified to take a given set and incorporate them into their next available batch for analysis. This one-shot approach to shipping samples was later revised to consist of three separate shipments, when concerns were raised as to whether the clients had sufficient storage facilities (e.g., cold room for storing dust-wipe samples) and whether the clients could ensure the integrity of the samples during storage. In addition, it was uncertain whether all clients would correctly retrieve the appropriate samples at the right time for analysis if they received all of the samples at one time.

Each shipment of proficiency-test samples was accompanied by copies of the Sample Tracking and Analysis Report Form and a cover letter. The Sample Tracking and Analysis Report Form, included in Appendix C, contained a list of PTS Sample IDs, information on the client receiving these samples, information on the laboratory to whom the client would provide these samples for analysis (only the laboratory name was specified; additional laboratory information was provided by the client), sample matrices and weights, and dates when the samples and results were shipped or received. Samples that were sent to different laboratories appeared on different copies of the form.

The recruitment staff provided the proficiency-testing service with necessary information for making sample shipments to clients (i.e., information to be placed on the Sample Tracking and Analysis Forms), including the clients' addresses, responsible parties, the laboratory(ies) to which a client would send samples, and the number and types of samples to send to each client. For ten clients, samples for double-blind (DB) Rounds 1, 2, and 3 were shipped on February 27, April 3, and June 1, 1998, respectively. These dates corresponded to approximately one, two, and four months following the date that the proficiency-testing service shipped proficiency-test samples to laboratories in Round 22 of the ELPAT Program. One of the 10 clients reported that they did not receive the first round of samples; a new batch of DB Round 1 samples was sent to this client on April 3, and DB Round 2 samples were sent on May 1. DB Round 1 samples for one client were sent on March 10. Samples for one client who enrolled late in the study were sent on April 21, May 18, and June 15. Samples for two clients who enrolled late in the study were sent on April 21, May 18, and June 1.

2.4.3 Receiving and Shipping of Samples by Clients

Upon being selected for participation in the study (and again with the batch of DB Round 1 samples), the clients were provided with written instructions and procedures for proper handling and shipping of the proficiency-test samples that they received in this study. These included the following:

- Samples should be stored in a locked area with limited access (e.g., cabinet, closet) until they are placed within the next available batch of field samples.
- Sample containers and their contents should not be opened or otherwise tampered with.
- Clients were to incorporate the samples randomly within their next available batch of field samples that are earmarked for analysis at that laboratory (or place the samples in their own batch if no field samples are available within one month of receiving the proficiency-test samples).
- Clients were to assign identifications (IDs) to all samples in a batch. For the proficiency-test samples, the clients were to remove the sample container from the outer plastic bag, noting the PTS Sample ID that is on this outer bag, and place a label on the sample container containing their assigned ID.

The instructions also contained detailed information on how to complete the Sample Tracking and Analysis Report Forms during the entire shipping and analysis process. As the proficiency-test samples were received by the client, the samples were given different IDs by the client, the samples were shipped to the laboratories, and the analytical results received from the laboratories. The clients were instructed to store these tracking forms in a secure location. See Appendix C for all instructions and procedures provided to the clients upon their enrollment in the study.

Once a client had incorporated the proficiency-test samples for a particular laboratory within its next available batch of regular field samples of the same type (dust wipe, paint chip), the client shipped the samples to the laboratory for analysis via its routine procedures, and ideally within one month of receiving the proficiency-test samples. The client was instructed to fax the tracking forms to the proficiency-testing service after recording the client sample IDs and the date that the samples were shipped to the laboratory in the tracking forms.

2.5 REPORTING RESULTS

When laboratory clients received proficiency-test samples analytical results from the laboratories, they transcribed the analysis results onto the Sample Tracking and Analysis Report Form, next to the corresponding sample IDs. The client then faxed the completed Sample Tracking and Analysis Report Form to the proficiency-testing service, along with any other pertinent information that the client felt might provide important insights about the analysis. After the proficiency-testing service had received all laboratory results that it expected within a given testing round, the results were entered into a spreadsheet and were forwarded for statistical summary and analysis. The proficiency-testing service also forwarded copies of the Sample Tracking and Analysis Report Forms and any accompanying ancillary information from the clients and laboratories. The proficiency-testing service also provided results of Round 22 of the ELPAT Program, which used proficiency-test samples created from the same batch as those used in the double-blind pilot study and which preceded the first round of the double-blind pilot study by approximately one month, along with the information from the ELPAT Program found in Table 2-1 for each type of proficiency-test sample. See Section 6.2 on details concerning types of data, data management, and data tracking issues.

3.0 STUDY PARTICIPANTS

This chapter documents the approaches taken to recruit risk assessors and lead-based paint inspectors for this pilot study who regularly submit dust-wipe and/or paint chip samples to NLLAP-recognized laboratories for analysis. As discussed in Chapter 2, the double-blind nature of this pilot study required that laboratory clients, rather than laboratories themselves, be recruited for the study.

As will be discussed in detail below, the recruitment process resulted in 19 clients being selected for participation in the study. Seventeen were privately-owned lead service providers, and two were state-operated health departments. Of the 19 clients selected, 12 provided analytical results (from 11 privately-owned laboratories and one state-operated health department's laboratory) on the proficiency-test samples they were provided.

Sections 3.1 and 3.2 present the general approaches used to perform recruitment and the criteria used to determine study eligibility. Section 3.3 documents the results of the recruitment process, including reasons why some potential clients did not wish to participate or were not selected for participation. Sections 3.4 and 3.5 present notes on client recruitment and communication throughout the course of the study.

3.1 RECRUITMENT APPROACHES

To recruit privately-owned risk assessors and lead-based paint inspectors, this study used two approaches:

- Publishing a news brief on the study in a newspaper widely-read by the lead hazard control industry, inviting interested parties to contact EPA.

- Making direct calls to lead-based paint inspectors and risk assessors.

Also, recruitment occurred at a meeting on the NLLAP held in December 1997, whose attendees included officials of several state-operated lead programs. Note, however, that it was not of interest to recruit many state-operated agencies for the study because of the potential that they might perform more QA/QC activities than private firms, resulting in more accurate laboratory results compared to the results associated with privately-owned risk assessors.

In both approaches to recruitment, recruiters interviewed potential study participants using the script found in Appendix B. In the interview, recruiters provided information on the double-blind pilot study, determined the client's interest in participating, and obtained information from interested clients to help assess whether a given client should be selected with higher or lower priority relative to other clients. Information obtained from interested clients included the following:

- whether they submit dust-wipe or paint chip samples on at least a monthly basis to one or more NLLAP-recognized laboratories

- names of NLLAP-recognized laboratories with which they contract, along with information on their testing volume, analytical methods, and detection limits

- approximate sizes of sample batches

- information on sample containers used and types of dust-wipes used.

Details on each approach to recruiting study participants and the outcome of implementing these approaches are as follows.

Recruitment Approach #1

In the first recruitment approach, a news brief on the pilot study was prepared and submitted to a trade journal for the lead hazard control industry for publication in an upcoming edition. This news brief, found in Figure 3-1, summarized the need for a double-blind proficiency testing program and how recruited clients of laboratories would assist with the program. Interested participants were asked to contact EPA or its contractor, Battelle.

A total of eight laboratory clients contacted EPA or Battelle as a result of reading the news brief, indicating their interest in participating. Seven of these clients were selected for the study.

Recruitment Approach #2

The second recruitment approach involved telephoning a sample of lead evaluation service providers (lead inspectors, risk assessors, and abatement contractors) who were included in the Lead Listing (the National Lead Service Providers' listing system) dated November 21, 1997. The service providers in the Lead Listing are grouped according to the state(s) in which they provide services. Service provider information in this list included name, telephone numbers, states in which the provider is certified as a lead inspector, and states in which the provider is certified as a risk assessor.[2]

Only service providers included within the New Jersey, Ohio, Pennsylvania, and Virginia segments of the Lead Listing were considered for recruitment in this study. This was done to

[2] The Lead Listing is operated by QuanTech, Inc., and ICF Information Technology, Inc., for the U.S. Department of Housing and Urban Development (HUD)'s Office of Lead Hazard Control, and is supported by the National Lead Assessment and Abatement Council (NLAAC). It can be accessed at *www.leadlisting.org* or by calling 1-888-LEADLIST.

Published in the "Newsline" column of Lead Detection and Abatement Contractor, January 1998 edition:

EPA Seeks Lab Clients for "Double Blind" Test Program

EPA is seeking clients of laboratories accredited by the National Lead Laboratory Accreditation Program (NLLAP) to participate in a double-blind proficiency testing pilot study.

At the present time, laboratories recognized by the NLLAP are required to successfully participate in the Environmental Lead Proficiency Analytical Testing Program (ELPAT). The ELPAT is a single-blind proficiency testing study. In a single-blind proficiency testing study, the participating laboratories are aware of when they receive proficiency-test samples. Because laboratories know when they are analyzing proficiency-test samples under the ELPAT, there is a potential to deviate from the routine analysis procedures provided for field samples, providing more attention to the analysis of the test samples. In a double-blind proficiency testing study, laboratory clients would submit proficiency-test samples to laboratories as routine field samples. Ideally, the test samples would be indistinguishable from the field samples. As laboratories would be unaware of when they would be analyzing proficiency-test samples, a double-blind study would give a more accurate evaluation of a laboratory's routine performance.

Clients selected to participate in the pilot study would be asked to insert proficiency-test samples in with batches of field samples that they are submitting for analysis to an NLLAP-recognized laboratory. A proficiency testing service would provide proficiency-test samples to the client at no charge. Laboratory clients interested in participating in the double-blind pilot study are asked to contact John Scalera of the EPA's Office of Pollution Prevention and Toxics by facsimile at 202/260-0001, or Robert Lordo of Battelle by phone at 614/424-4516 by Jan. 20.

Figure 3-1. News Brief on the Double-Blind Proficiency Testing Pilot Study

improve the overall recruitment success rate, as these four states were among those that required service providers to use NLLAP-recognized laboratories at the time of this pilot study.

Telephone recruitment of laboratory clients in the Lead Listing occurred from January through April, 1998, with most of the recruitment completed before March, 1998. A total of 55 service providers were contacted. Results of these contacts were as follows:

- 21 contacts indicated an interest in participating. (As discussed in Section 3.3, not all of these contacts were found to be eligible for participation, and two contacts later declined interest.)

- 10 contacts indicated that they were not interested in participating, for reasons such as a lack of sufficient dust and/or paint sampling for lead testing, they were too busy, the study would require too much effort, or they did not have the resources to participate.

- 24 providers either could not be reached, did not reply to messages left by recruiters, were no longer in the lead evaluation business, or no longer responded after requesting time to review study information provided by the recruiters.

Some of the 21 interested contacts became interested only when a policy was adopted in this study to reimburse study participants for analytical costs associated with the proficiency-test samples. The reimbursement issue is discussed further in Section 3.4. Of the 21 interested clients, 11 were accepted into the study (with one client dropping out prior to the start).

Other Recruitment Approaches

The EPA Work Assignment Manager discussed the double-blind pilot study at a meeting held on December 17, 1997, where state and federal government representatives interested in environmental-lead testing met to discuss NLLAP-related subjects. At this meeting, five representatives of state health departments (from five different states) who collected dust-wipe and/or paint chip samples in household risk assessments and had the samples analyzed for lead by NLLAP-recognized laboratories indicated an interest in participating in this pilot study. Once all recruitment efforts in this study were completed, two states remained interested and were selected to participate in the study.

When initially formulating the idea of a double-blind pilot study and discussing it with various representatives of the lead inspection industry, the EPA Work Assignment Manager identified two lead inspectors who were interested in participating. However, these two inspectors were not available to participate once the study was ready to begin.

3.2 CLIENT ELIGIBILITY

During the initial telephone contact with a potential study participant, responses to the questions in the telephone script (Appendix B) were used to determine if the laboratory client was eligible for the pilot study and to prioritize those clients found to be eligible. To participate in the pilot study, a client recruited via telephone must have met the following criteria regarding dust and/or paint sample collection and analysis:

- Contract with an NLLAP-recognized laboratory to analyze the given sample type - The client must have routinely submitted dust and/or paint samples to NLLAP-recognized laboratories for lead testing. This criterion allowed an evaluation of the current routine performance of a subset of NLLAP laboratories.

- Use the PaceWipe™ or similar towelettes to collect dust samples (only for clients recruited to submit dust-wipe samples) - All proficiency-test dust-wipe samples in this study consisted of the PaceWipe™. Because of the need to have proficiency-test samples indistinguishable from regular field samples, it was necessary to enroll clients who collected dust samples in the field using the PaceWipe™ or a towelette that is similar in appearance to the PaceWipe™.

Other criteria that were used to prioritize clients for selection in this pilot study but were not absolutely necessary for the clients to meet included the following:

- Submit batches of the given sample type at least monthly - Because of the desire to have all samples for this pilot study analyzed between scheduled rounds of the ELPAT Program, it was desired to enroll clients who submitted dust-wipe and/or paint chip samples to an NLLAP-recognized laboratory at least once per month. This enabled the double-blind samples to be sent to the laboratories within one month after the clients received them. While clients met this criterion during enrollment, some occasionally did not have the field samples to ship to a laboratory in a given month, and therefore, submitted only the proficiency-test samples in that round.

- Do not perform a double-blind procedure on their contracted laboratories, or do not contract with laboratories in the HUD Grantee program - Laboratories involved in double-blind testing either through the efforts of their clients or through their participation in HUD's Lead-Based Paint Hazard Control Program (HUD Grantee) may have been already aware that they were being evaluated via double-blind procedures. This may enhance their routine performance relative to laboratories that participate only in single-blind proficiency testing. Therefore, higher priority was placed on clients who did not perform double-blind procedures or who contracted with laboratories that were not affiliated with the HUD Grantee program.

- Contract with larger laboratories - Initially, clients who contracted with NLLAP-recognized laboratories with high testing volumes were to have a higher priority for selection than clients who used exclusively small laboratories. As large laboratories tended to do the majority of sample analyses, this pilot study aimed to have more proficiency-test sample analyses performed by large laboratories compared to small laboratories. However, in the final selection process, it was desired to have as many different laboratories involved as possible, and so this criterion was relaxed.

- Use centrifuge tubes for shipping dust-wipe samples to the laboratory – A dust sample result was reported as a total lead amount in the sample. To ensure that a laboratory would not cause a quantitative loss of lead or dust when removing a dust-wipe sample from its shipping container, clients receiving dust-wipe proficiency-test samples needed to ship such samples in centrifuge tubes, rather than plastic bags. As paint chip samples were re-weighed prior to analysis and their results reported as a percentage of lead by weight, use of plastic bags for shipping paint chip samples was deemed acceptable.

In addition, to protect any one laboratory from being over-represented in this study, some clients enrolled later in the study were prioritized based on the NLLAP-recognized laboratory(ies) to which they contract. Clients who contracted with a laboratory that was considered sufficiently represented by other clients enrolled earlier were occasionally given a low priority for selection, while those contracting with a laboratory not yet represented in the study were given a high priority.

3.3 CLIENT SELECTION AND ENROLLMENT

As mentioned in Section 3.1, 29 privately-owned laboratory clients expressed an interest in participating in this pilot study, either through responding to the news brief or being contacted by the recruitment staff. Once these clients were identified, the study team used the criteria mentioned in Section 3.2 to select those clients who would participate in the study, determined which type(s) of proficiency-test samples (dust-wipes, paint chips) each selected client would be submitting in the study, and determined to which laboratory(ies) the clients would submit these samples.

Of the 29 privately-owned clients who showed an interest in participating, 11 were not selected for the study for the following reasons:

- Before receiving study participation instructions and proficiency-test samples, 3 clients decided that they did not want to participate.
- 3 did not submit dust or paint samples to an NLLAP-recognized laboratory at least once per month.
- 2 used baby wipes rather than towelettes to collect dust samples and did not collect paint samples.
- 2 did not provide sufficient information to allow them to be considered.
- 1 used baby wipes rather than towelettes to collect dust samples, but did collect paint samples. However, they contracted paint sample analysis with a laboratory which was already considered to have sufficient participation in this study by the time this client was interviewed.

The remaining 18 privately-owned clients, in addition to two state health departments, were selected to participate in the study. One of the privately-owned clients dropped out of the study prior to its start (due to the need to redirect staff to other work), for a total of 19 clients. Based on information that these clients provided on their dust and paint sampling volume, 11 were selected to submit both dust and paint proficiency-testing samples for analysis in this study, 2 were to submit only dust samples, and 6 were to submit only paint samples.

Table 3-1 documents the 19 clients successfully recruited for this pilot study, the type(s) of proficiency-test samples that they would submit in each testing round, and the NLLAP-recognized laboratory(ies) to which they would submit the samples. For this report, the clients are identified according to a unique alphabetic ID, while the laboratories are identified by a unique two-digit numeric ID. From this table, the following should be noted:

- The 19 clients were associated with 16 different NLLAP-recognized laboratories.
- 7 clients were to submit samples to two different laboratories.
- 2 laboratories were to have received samples from three different clients, and 5 laboratories were to have received samples from two different clients.

Perhaps of most importance, as indicated by the footnotes to Table 3-1, 7 of the 19 clients (clients M through S) decided not to participate once the pilot study was begun, by either not providing sample containers or not reporting back the results of the proficiency-test sample analyses in any testing round. Therefore, only 12 clients (clients A through L) actually participated in the pilot study, representing 12 laboratories (laboratories 01 through 12). The issue of client non-response during the study is discussed in Section 3.5.

All but one (laboratory 08) of the 12 laboratories who analyzed proficiency-test samples in the double-blind pilot study were NLLAP-recognized at the time that they analyzed proficiency-test samples in Round 22 of the ELPAT Program. Thus, they were among the 118 reference laboratories whose results determined the target lead amounts associated with the proficiency-test samples in Round 22. For each of the three proficiency-test sample types that were included in the double-blind pilot study, Figure 3-2 contains a bar chart of the observed distribution of the ELPAT Round 22 results for the 118 laboratories plus laboratory 08, with results for the laboratories participating in the double-blind pilot study highlighted in black within the bars. These charts show that while most of the ELPAT Round 22 data for the double-blind pilot study laboratories were in the upper half of the distribution of reference laboratory data for each sample type, the data for the double-blind laboratories were good representations of the reference laboratory data. This was confirmed when statistical comparisons were made between ELPAT Round 22 data for the double-blind laboratories and ELPAT Round 22 data for the remaining NLLAP-recognized laboratories. When both a two-sample t-test and the Mann-Whitney nonparametric test were applied to these data, no significant differences were observed between these two groups of laboratories at the 0.05 level for each of the three sample types.

Once the 19 clients in Table 3-1 were selected for the pilot study, letters of acceptance were prepared and sent to them. These letters were accompanied by an attachment providing

Table 3-1. Clients Recruited to Participate in This Pilot Study, the Type(s) of Proficiency-Test Samples That They Would Submit in Each Testing Round, and the NLLAP-Recognized Laboratory(ies) To Which They Would Submit the Samples

Client ID	Number and Type of Proficiency-Test Samples Submitted in Each DB Testing Round	Laboratory ID
A	2 dust samples	10
	2 dust, 4 paint samples	05
B	2 dust samples	08
C[1]	2 dust, 2 paint samples	06
	2 dust, 2 paint samples	15
D	2 dust, 2 paint samples	03
	2 dust, 2 paint samples	04
E	2 dust, 2 paint samples	04
F	2 dust samples	08
G	2 dust, 2 paint samples	02
	2 dust, 2 paint samples	09
H	4 dust, 4 paint samples	02
I	2 dust, 2 paint samples	01
	2 dust, 2 paint samples	07
J	4 paint samples	08
K	2 paint samples	11
L	4 paint samples	12
The following clients were recruited but did not participate once the study begun		
M[2]	*2 dust, 2 paint samples*	*06*
	2 dust, 2 paint samples	*13*
N[2]	*2 dust, 2 paint samples*	*04*
	2 dust, 2 paint samples	*14*
O[2]	*4 dust, 4 paint samples*	*not determined*
P[3]	*2 dust, 4 paint samples*	*09*
Q[2]	*2 paint samples*	*11*
R[3]	*2 paint samples*	*16*
S[2]	*2 paint samples*	*16*

[1] This client ended its contract with Laboratory 15 prior to the start of the study, and any samples earmarked to be sent to Laboratory 15 were sent to Laboratory 06 instead.
[2] Withdrew from the study as they did not submit sample containers to the proficiency-testing service.
[3] Withdrew from the study as they did not report analytical results on the proficiency-test samples provided to them, despite several attempts to contact them for the results.

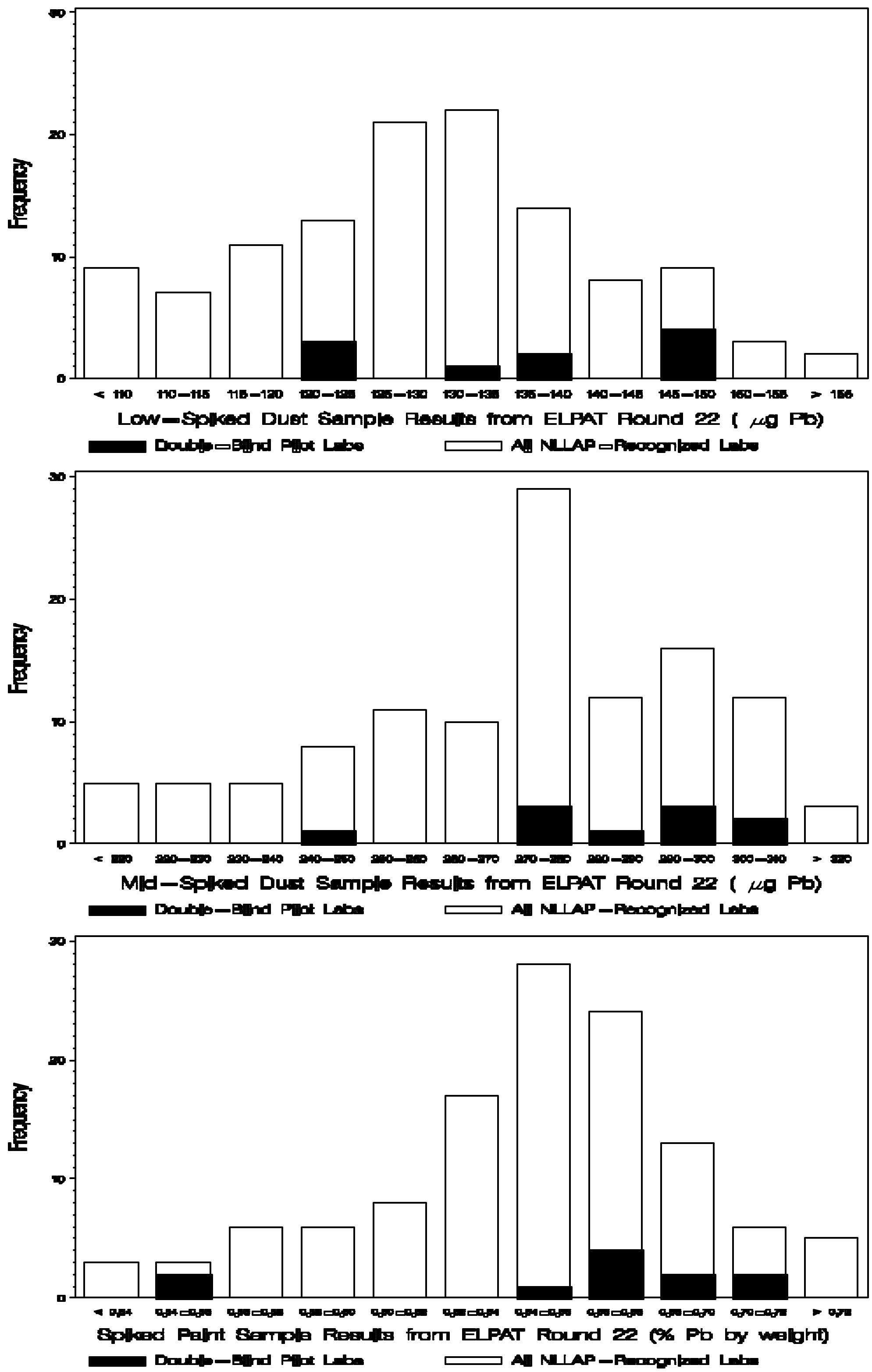

Figure 3-2. Histograms of the Proficiency-Test Sample Analysis Results from Round 22 of the ELPAT Program for 118 NLLAP-Recognized ("Reference") Laboratories, with the Results for the 12 Laboratories Participating in the Double-Blind Pilot Study Highlighted

details on the pilot study which the participating clients needed to know (such as objectives and point of contact), the tasks which the clients would perform, and other necessary instructions. The letters were sent early enough to give clients sufficient time to submit sample containers to the proficiency-testing service, so that they had sufficient time to transfer the proficiency-test samples for the first DB testing round to these containers and to ship these samples to the clients within the specified time schedule. The generic form of the acceptance letter and the attachment are found in Appendix C.

Once the acceptance letters were sent, the proficiency-testing service was supplied with the following information for each selected client: name, address for shipping the proficiency-test samples, and the name(s) of the laboratory(ies) to receive these samples from the client. At this point, the proficiency-testing service became the primary point of contact with these clients. The proficiency-testing service recorded this information on copies of the Sample Tracking and Analysis Report Form (Section 2.4.2).

3.4 RECRUITMENT EXPERIENCES

The following issues and experiences arose while conducting the recruitment process of laboratory clients in this pilot study:

"What's in it for me?"

When initially contacted, some laboratory clients did not recognize how a double-blind program would allow them to have greater confidence in the accuracy of the lead levels being reported back from the laboratories on the samples that they collect. As a result, they either did not have a desire to participate due to the responsibilities involved, or wished to be reimbursed in some way for the resulting expense.

Need for reimbursement

When the recruitment effort began, it was apparent that the interest of clients to participate in this study would increase if they were reimbursed for the analytical costs of the proficiency-test samples that were provided to them. While this was particularly true for small organizations who had limited cash flow, some larger organizations stipulated that compensation for these costs was a requirement for their participation. Therefore, to improve the success rate of the recruitment effort, the study team decided to adopt a policy of reimbursing the clients for costs billed to the client by the laboratory to analyze the proficiency-test samples. In order to receive this reimbursement, clients had to provide copies of the laboratory's invoice, showing either the cost per sample or the number of samples analyzed and the total analysis cost material, thereby documenting that the proficiency-test samples were analyzed. Clients would not be reimbursed for other costs to participate in the study, such as shipping costs and costs associated with providing sample containers to the study.

While the decision to reimburse clients increased the rate of interest in the study, the issue of whether compensation is a viable part of an actual double-blind program needs to be addressed (e.g., what would be the source of the compensation money). In addition, providing compensation would require a support staff, and therefore, additional resources.

Difficulty in making a contact

It was occasionally difficult to make contact with a potential study participant based on just the name and telephone number information. For example, frequent changes in a client's name and telephone number from that specified in the Lead Listing sometimes made it difficult to reach some clients.

Recruiters were often intercepted by answering machines or voice mail when making calls. When someone did answer the telephone, this person was not always knowledgeable on NLLAP issues, or the most appropriate contact was occasionally not available (e.g., was in the field doing sample collection). In these situations, recruitment staff left telephone messages, describing the study and requesting an appropriate contact to call back. However, in many cases, the client did not call back.

Desire to review responsibilities

Some clients wished to have materials on the study sent to them for review prior to giving notification that they were interested in participating. The materials prepared for faxing to these clients included a copy of the news brief in Figure 3-1 and selected materials in the attachment to the acceptance letter found in Appendix C. Discussions with others at the client's headquarters were occasionally required before a client indicated a desire to participate.

Limited work involving environmental sampling for lead

Some clients did not feel that they could participate due to the small number of dust and/or paint chip samples that they submit to an NLLAP-recognized laboratory in a given month. This should not be a major concern if the client is asked to include only two proficiency-test samples in a given batch. In fact, there is one issue that supports having proficiency-test samples placed in small batches. If the proficiency-test sample results are considerably different than those for other samples within the batch (e.g., are very high), then some laboratories may suspect that the samples are some kind of reference material. This is more likely to happen in large batches than in small batches, especially if the lead content is consistent from sample to sample within a batch.

Use of an appropriate dust wipe

While many potentially-interested clients used a pre-moistened wipe to collect dust-wipe samples, it is important in a double-blind program that this wipe closely resemble the type of wipe used in the proficiency-test samples. As mentioned earlier, the ELPAT Program (and, therefore, this pilot study) uses the PaceWipe™, which resembles a small towelette like those stored in individual wrappers and provided to customers in many fast-food restaurants. However, several clients use baby wipes rather than the PaceWipe™ or a towelette. Baby wipes are generally larger, thicker, and have a different consistency and aroma than the PaceWipe™ or a towelette. Therefore, to limit the likelihood that a laboratory would distinguish a difference between the proficiency-test samples and regular field samples on the basis of the type of wipe used, interested clients who used baby wipes were not considered for submitting dust proficiency-test samples in this pilot study.

Need to provide materials for selected clients

As discussed in Section 2.4.1, clients were asked to provide sample containers to ensure that the proficiency-test samples would be placed in the same containers that clients used to store field samples. However, two clients in this study were provided default sample containers (centrifuge tubes that are different from the containers used in the ELPAT Program) to use in this study in order for them to agree to participate in the study. In this instance, the clients were also supplied with a sufficient number of empty containers for all of their regular field samples for the batch in which the proficiency-test samples would be placed.

In addition, to ensure participation, the study provided PaceWipes™ to one client to use in collecting field samples that would be included with the dust proficiency-test samples.

3.5 INTERACTION WITH CLIENTS DURING THE STUDY

During the pilot study, the proficiency-testing service was the primary point of contact with the participating laboratory clients. It was the proficiency-testing service's responsibility to ensure that the clients received their proficiency-test samples on time within each testing round, to answer any questions that the clients had during the study, and to collect analytical results on the proficiency-test samples from the clients within each testing round. The following are issues that were encountered when interacting with the clients during the course of the study.

Some clients did not provide information/materials on time or at all

At the start of the study, some clients did not meet the prescribed deadline for submitting clean, empty sample containers to the proficiency-testing service. During the study, some clients did not report their results back to the proficiency-testing service in a given DB testing round within the one-month period after receiving the proficiency-test samples. Client delays in reporting results or materials were due to a number of reasons, such as when clients had not yet collected a sufficient number of field samples to accompany the proficiency-test samples within a batch, or when clients simply needed reminding. In these situations, it was necessary for the proficiency-testing service to follow-up with the clients to check on the status of the materials or information and reasons for delay. As was observed in the recruitment process, these follow-up calls frequently resulted in a request to call back, either by reaching voice mail or a secretary/ receptionist. This delayed the receipt of materials or information even more.

As of the scheduled release of the DB first-round proficiency-test samples to the clients, six clients had not provided the proficiency-testing service with sample containers. (These clients were not scheduled to receive default containers.) As seen in Table 3-1, five of these clients never provided containers, leading to their removal from the study.

Despite several attempts to contact them, the proficiency-testing service was unable to determine whether two clients submitted proficiency-test samples for analysis in any DB testing round and that they received analytical results back from the laboratory. Therefore, as seen in Table 3-1, these clients were removed from the study.

In one instance, after repeatedly trying to get results from one client for a given DB testing round, the client finally called and read the results to the proficiency-testing service over the telephone, rather than by providing a completed Sample Analysis and Tracking Form. However, as the sample IDs that the client provided over the telephone did not match the IDs that the proficiency-testing service had for these samples, these data were considered unreliable and were not used in the final analysis.

<u>Occasional shipping of exclusively proficiency-test samples for analysis</u>

In situations where it was uncertain whether certain clients would be able to create a batch of regular field samples into which the proficiency-test samples could be placed, clients were instructed to submit the proficiency-test samples within a batch by themselves and submit the batch to the laboratory. Here, the samples would still need to be disguised by the client as field samples by specifying fabricated information on where and how the sample was collected in the field.

In one situation occurring in DB Round 1, a client had no field samples to place in a batch. Therefore, the client added blanks to the batch to help disguise the batch from containing exclusively proficiency-test samples. However, the client asked to be reimbursed for the analysis of all samples in the batch, including the blanks. The pilot study granted this request. The client also indicated that in DB Rounds 2 and 3, the proficiency-test samples would be labeled as “follow-up” samples to the DB Round 1 batch.

<u>Non-receipt of samples</u>

In two instances, a client reported that they did not receive a batch of proficiency-test samples for a given testing round. This occurred for one client in DB Round 1 and a different client in DB Round 3. For the DB Round 1 situation, sufficient samples were prepared (and sample containers acquired from the clients) to allow for a second batch to be provided to the client for that round. This required, however, an extension on the amount of time that a client had to report results back to the proficiency-testing service. Because the selected period for sample analysis and reporting was nearly over, no replacement samples were shipped to the client who reported no Round 3 samples were received.

Reported measurement units associated with sample results

As indicated in the instructions to the clients, the proficiency-test samples were to be accompanied to the laboratory with fabricated information on where the sample was “collected” within some residence. Some clients also provided fabricated sampling areas (e.g., 0.5 ft^2) associated with the “sample.” This practice caused considerable confusion on when a lead measurement being reported on the Sample Tracking and Analysis Forms represented a total lead amount in the sample or a lead amount per unit of sampling area. When clients specified lead measurement as lead per unit area rather than as total lead amount, it was necessary for the proficiency-testing service to follow-up with these clients to verify the sample areas that were specified to the laboratory. Often, several calls were necessary to get the needed information. Also in this pilot study, a client occasionally reported the wrong units on the Sample Tracking and Analysis Form, such as reporting a result as micrograms when it actually represented micrograms per square foot. Such errors were found at the end of the study when the proficiency-testing service obtained copies of the laboratory report forms from the clients or laboratories to investigate the validity of certain unusual data values. Proper reporting of measurement units is one example on how using the client as a “middleman” in reporting laboratory results can increase the likelihood that results will be reported inaccurately to the proficiency-testing service.

4.0 DATA ANALYSIS PROCEDURES

As presented in Section 1.3, the primary objectives of this double-blind pilot study were to

- Design and test a protocol for double-blind proficiency testing under the NLLAP, which involves evaluating logistical issues, identifying responsibilities and investigating minimum performance standards for laboratories

- Compare analytical performance of NLLAP-recognized laboratories between the double-blind pilot study and single-blind testing (i.e., Round 22 of the ELPAT Program).

Accomplishing both of these objectives required statistical analysis of the data collected during the pilot study. The corresponding statistical objectives were to

- Summarize the measured lead levels for each laboratory and across all laboratories

- Perform similar summaries on lead levels expressed relative to the ELPAT Round 22 target levels (Table 2-1), and to compare a laboratory's analytical results with what the laboratory reported in ELPAT Round 22.

- Assess quantitative information (e.g., variability characterization) needed to determine appropriate statistical proficiency criteria for a double-blind program.

The methods used to prepare the data for analysis and the analysis methods used to accomplish the statistical objectives are described in the sections that follow. Results of implementing these methods through use of the SAS® System on the collected data in this study are provided in Chapter 5.

4.1 DATA ANALYSIS ENDPOINTS AND DATA HANDLING

Once the proficiency-testing service provided the pilot study data in electronic format, these data were placed into SAS® datasets for statistical summary and analysis. Prior to beginning the analysis, however, the data in these datasets were hand-checked for accuracy against the completed Sample Tracking and Analysis Report Forms (also provided by the proficiency-testing service) as discussed in Section 6.3, and any necessary corrections were made to the datasets.

As discussed in Section 6.1, a common batch of bulk source material was used to generate each proficiency-test sample of a given type in both the double-blind pilot study and Round 22 of the single-blind ELPAT Program. Therefore, if the samples of a given type were homogenous and laboratories are proficient, the measured lead amounts should be approximately the same across samples, regardless of which laboratory performed the analysis, and should be approximately equal to the target lead amounts as reported in Round 22 of the ELPAT Program. Therefore, data summary and analysis focused on the following two endpoints, whose values were determined for each proficiency-test sample:

- The measured lead amount in the sample, as reported by the laboratory analyzing the sample and forwarded by the client who was provided the sample (expressed in micrograms of lead for dust-wipe samples and percent lead by weight for paint-chip samples).

- Percent recovery associated with the sample, equal to the measured lead amount divided by the target lead amount associated with the bulk source material, as determined in Round 22 of the ELPAT Program (Table 2-1), multiplied by 100%.

These data were categorized according to testing round, client, and laboratory.

4.2 IDENTIFYING AND HANDLING STATISTICAL OUTLIERS

Prior to statistical summary and analysis, the data were investigated for values that appeared to be extreme (e.g., very high or very low) compared to other values for proficiency-test samples that were created from the same batch of bulk materials. Initially, this was done by plotting the data and noting those data points whose values were considerably different from the others for the given sample type. These data points were reported to the proficiency-test service, who attempted to contact the appropriate clients to determine the cause of the extreme values (e.g., reporting errors, mislabeled samples, invalid analysis). In some cases, if the client contact could not completely address the validity of the extreme data values, the proficiency-test service contacted the laboratories (once all laboratory testing was completed).

While client and/or laboratory contacts occasionally resulted in corrections to certain extreme data values, other values were found to be legitimate as reported for purposes of this analysis. However, their presence can overly inflate variability estimates within the statistical analyses. Therefore, a statistical procedure was used to identify "statistical outliers," or those data values identified as unusually high or low according to some pre-specified statistical criteria, and the statistical summaries and analyses were performed both with and without the statistical outliers included. The "generalized extreme-Studentized deviate (ESD) many-outlier" procedure documented in Rosner (1983) was applied independently to both the untransformed and log-transformed lead amounts for each of the three sample types (low dust, mid dust, paint), with data for all three testing rounds combined. Within each application of the outlier detection procedure, up to 10 outliers could be identified at an overall significance level of 0.05 (i.e., the probability of identifying more outliers than are actually present is no higher than 0.05). This procedure assumes that the data have either a normal or lognormal distribution, which has historically been observed in previous studies.

Section 5.1 contains those reported lead amounts that were identified as being statistical outliers according to the procedure by Rosner (1983).

4.3 SUMMARY OF REPORTED LEAD AMOUNTS

To begin to see how well the laboratories performed relative to the target amount and to each other, simple summary statistics of the reported lead amounts and percent recoveries were generated. Note that these initial summaries were meant primarily to document the reported data and to provide preliminary, observational information on data variability and on overall data accuracy relative to target levels from ELPAT Round 22. Statistical comparisons to evaluate whether differences in variability from one testing round to the next were significant were made in subsequent sections of Chapter 5, using methods documented in Section 4.4 below.

Initially, the data were portrayed graphically by sample type and testing round to provide initial information on how the data varied across laboratories (due to multiple laboratories analyzing the same type of sample) and within laboratories (due to each laboratory analyzing at least two samples of a given type for the client). In addition, the graphs showed how lab-to-lab variation differed among the double-blind testing rounds as well as with the variability in data from Round 22 of the ELPAT Program for the same laboratories and sample batches.

The following types of tabular summaries of the above two endpoints were prepared for each proficiency-test sample type:

- Summaries of individual sample results, performed by double-blind testing round across all samples, regardless of client or laboratory

- Summaries of individual sample results, performed by double-blind testing round and laboratory

- Summaries of the average sample result for a client/laboratory combination, performed by double-blind testing round across all such combinations.

The tabular summaries specified by the first and third bullets included the number of results, arithmetic mean, standard deviation, minimum value, maximum value, and quartiles (i.e., 25^{th}, 50^{th}, and 75^{th} percentiles). The summaries specified by the second bullet included the number of results, arithmetic mean, and standard error of the mean.

For comparison purposes, the tabular summaries specified by the first bullet above also included summaries of data from Round 22 of the ELPAT Program. Three subsets of ELPAT Round 22 data were summarized: 1) data for only those laboratories participating in the double-blind pilot study, 2) data (non-Winsorized) for the 118 NLLAP-recognized laboratories participating in ELPAT Round 22, and 3) data for all laboratories participating in ELPAT Round 22. These summaries were included to allow initial comparisons of various types between the double-blind pilot study data and ELPAT Round 22 data for a given proficiency-test sample type.

Tabular data summaries were generated using the UNIVARIATE procedure in the SAS® System. In addition, summaries of the double-blind pilot study data were calculated both including and excluding the statistical outliers identified by the methods in Section 4.2.

4.4 CHARACTERIZING COMPONENTS OF VARIATION

Statistical criteria for evaluating a laboratory's proficiency within a double-blind program can be established only after characterizing the different sources of data variation that are present in the proficiency-test sample analysis results. Variability that needed to be considered in the single-blind ELPAT Program (e.g., lab-to-lab variability) must also be considered in a double-blind program. However, a double-blind program may need to consider additional sources of variability, such as variability due to analyzing samples originating from different clients and due to laboratories being unaware of when a sample they are analyzing is a proficiency-test sample. Furthermore, as laboratories may test multiple samples of a given type in a double-blind testing round, within-laboratory variability can also be characterized.

The initial design of the double-blind pilot study permitted up to three sources of variation in the analytical results to be measured:

- lab-to-lab variation (due to the presence of multiple laboratories)
- client-to-client variation within a laboratory (due to multiple clients sending samples to the same laboratory)
- within-client and laboratory variation (due to laboratories receiving at least two replicate proficiency-test samples from a client within a testing round)

Furthermore, the study included multiple double-blind testing rounds to investigate how these sources of variation contribute differently to total variability in the reported results from one testing round to another.

A lack of sufficient data prohibited client-to-client variation within a laboratory (i.e., the second bullet in the previous paragraph) to be characterized appropriately within this pilot study. Therefore, the statistical analysis only took into consideration two components of total variability in the double-blind pilot study data: <u>lab-to-lab variability and within-lab variability</u>. Within-lab variability, therefore, actually represents variability in results for different clients as well as the same client within a common laboratory in a given testing round.

Analysis of variance procedures were used to characterize lab-to-lab and within-lab variability in the double-blind pilot study data for each DB testing round. The data analyzed in these procedures were <u>log-transformed results for dust-wipe samples and untransformed paint sample results</u>. A log transformation was made to the dust-wipe data as these data represented a total measured lead amount in the sample. Such measurements have been found to originate more likely from a lognormal distribution than a normal distribution (i.e., majority of low values, with an extended right tail to represent occasional high values). Paint sample results were not log-transformed, as these measurements were expressed as percentages (i.e., percent lead by sample weight), and typically, percentages do not require a logarithmic transformation to achieve assumptions of a normal distribution more accurately.

4.4.1 The Full Statistical Model

To characterize total variability in the double-blind pilot study data, the contribution of lab-to-lab variability to this total variability, and how the variability changes across the testing rounds, a mixed-effects model was developed. This model had the following form:

$$Y_{ijk} = m + R_j + w_{i(j)} + e_{k(ij)}, \qquad (1)$$

where Y_{ijk} is the measured lead amount (log-transformed for dust samples, untransformed for paint samples) for the kth sample that the ith laboratory analyzed in the jth DB testing round (j=1, 2, 3);
µ is the overall mean response across all samples, laboratories, and DB testing rounds;
R_j is the (fixed) effect of the jth DB testing round;
ù$_{i(j)}$ is the (random) effect of the ith laboratory within the jth DB testing round; and
g$_{k(ij)}$ is random error associated with the kth sample analyzed by the ith laboratory in the jth testing round (i.e., that portion of the measured lead amount that is not explained by the model).

Furthermore, the effect ù$_{i(j)}$ was assumed to be a random variable having a normal distribution with mean zero and variance L_j^2 (j=1, 2, 3). This variance L_j^2 is a measure of lab-to-lab variability within a DB testing round and is allowed to differ across testing rounds. The effect g$_{k(ij)}$ was assumed to be a random variable having a normal distribution with mean zero and variance ó$_j^2$ (j=1, 2, 3). This variance ó$_j^2$ is a measure of within-lab variability within a DB testing round and is also allowed to differ across testing rounds.

Therefore, within a given DB testing round, model (1) assumes that a given measurement is equal to an overall predicted value (given by µ+R_j), plus or minus some random amount that comes from a normal distribution with mean zero and variance L_j^2+ó$_j^2$. The variability associated with this additional random amount has components corresponding to lab-to-lab (L_j^2) variability and within-lab (ó$_j^2$) variability measures. Model (1) is called a “full” model as the overall predicted value and the two variance components can differ from one DB testing round to another.

Model (1) was fitted, and the variances of each component estimated, using restricted maximum likelihood routines within the MIXED procedure in the SAS® System. For each sample type, two fittings of the model were made: one including and one excluding the statistical outliers identified by the methods in Section 4.2. With three sample types in the double-blind pilot study, model (1) was fitted a total of six times.

4.4.2 Reduced Forms of the Full Statistical Model

The key feature of model (1) was that it allowed both the lab-to-lab variability and within-lab variability to differ from one DB testing round to another. To determine whether lab-to-lab variability differed significantly across DB testing rounds (and also whether within-lab variability differed significantly across rounds), a likelihood ratio test approach was taken. This approach involved fitting reduced versions of model (1) to the same set of double-blind pilot study data that were used in fitting model (1), where the reduced models did <u>not</u> allow one or both of the variance components to differ in their estimates from one round to another. Three different reduced models were considered:

1. The lab-to-lab variability estimate did not differ from round to round, but the within-lab variability estimate could (i.e., $L_1^2=L_2^2=L_3^2$, but $\acute{o}_j^2$ (j=1, 2, 3) could differ).

2. The within-lab variability estimate did not differ from round to round, but the lab-to-lab variability estimate could (i.e., $\acute{o}_1^2=\acute{o}_2^2=\acute{o}_3^2$, but L_j^2 (j=1, 2, 3) could differ).

3. Both lab-to-lab variability and within-lab variability estimates were consistent across testing rounds (i.e., $L_1^2=L_2^2=L_3^2$ and $\acute{o}_1^2=\acute{o}_2^2=\acute{o}_3^2$).

For model (1) and for each of the three reduced models, the estimate of the likelihood function was obtained from the model fit. Then, the likelihood estimates for two specific models, where one model was reduced from the other, were used in a likelihood ratio test to determine whether the reduced model was associated with a statistically worse fit compared to the other. This test determined whether the additional model parameters associated with the more complete model (e.g., different variability estimates for random effects between testing rounds) were statistically necessary.

To describe the likelihood ratio test approach, consider the example that tests whether or not both $L_1^2=L_2^2=L_3^2$ and $\acute{o}_1^2=\acute{o}_2^2=\acute{o}_3^2$ in model (1). In this test, the ratio of the likelihood for model (1) to the likelihood for reduced model #3 above was calculated. The test statistic T, equal to -2 times the log-transformed likelihood ratio, was calculated and compared to the 95th percentile of a chi-square distribution with degrees of freedom equal to the difference in the number of model parameters between the two models (4 in this example). If T exceeded this percentile, then it was concluded that the fit associated with the full model was significantly better (at the 0.05 level) than the fit associated with reduced model #3 (and, therefore, that some difference in variability estimates exists either for lab-to-lab or within-lab variability). Other likelihood ratio tests that compare model (1) with reduced models #1 and #2 specifically test for equal variability estimates between DB testing rounds for lab-to-lab variability and within-lab variability, respectively.

As with model (1), the likelihood ratio approach was applied separately by proficiency-test sample type and whether or not statistical outliers (as identified through the procedures documented in Section 4.2) were included in the analysis.

4.4.3 Comparisons of Lab-to-Lab Variability Between the Pilot Study and ELPAT Round 22

While the likelihood ratio test approach discussed in the previous subsection tested for significant differences in lab-to-lab variability and in within-lab variability between the three DB testing rounds, it was also of interest to test whether lab-to-lab variability in the DB pilot study differed significantly from the variability across laboratories in Round 22 of the ELPAT Program for the same proficiency-test sample type. This objective could not be addressed by adding ELPAT Round 22 data to the double-blind data in the above statistical modeling analyses, as each laboratory analyzed only one proficiency-test sample of a given type in ELPAT Round 22, thereby preventing within-lab variability from being estimated in the statistical modeling procedures for ELPAT Round 22. Instead, the following approach was performed (on log-transformed lead amounts in dust-wipe samples and on untransformed paint-lead measures given as percent lead by weight):

1. Arithmetic averages were calculated of the reported results for each laboratory within each DB testing round (ignoring whether the results were associated with different clients).

2. ELPAT Round 22 results were identified for the same group of laboratories.

3. Levene's test of homogeneity of variance (Snedecor and Cochran, 1989) was applied to the combined set of measures in #1 and #2 to test whether variability in these measures differed between the four testing rounds (i.e., the three DB testing rounds and ELPAT Round 22).

Because a laboratory's result from ELPAT Round 22 was for a single sample, and its result in a given DB testing round was an average of multiple samples (where the number of samples can differ from one laboratory to another), the data used in this analysis for a given laboratory were not generated from a source having equal variability across all four testing rounds. This is a statistical concern that must be recognized when interpreting the results of this analysis.

The analysis was applied separately to each proficiency-test sample type and whether or not statistical outliers from the double-blind study (as identified through the procedures documented in Section 4.2) were included. Results of this analysis are presented in Section 5.4.

A similar application of Levene's test of homogeneity of variance was performed on the individual sample results from the three double-blind testing rounds and from Round 22 of the ELPAT Program for the laboratories participating in the double-blind pilot study, in order to evaluate differences in total variability of individual sample results across these four testing rounds. In this application, results in each testing round were analyzed without regard to the laboratory which generated them, in order to obtain a general idea of whether double-blind testing data have a different underlying variability compared to single-blind data from the ELPAT Program. Note that this is an approximate test, as the laboratory effect on these data within a testing round was ignored. Results of this statistical test are presented in Section 5.2.

4.4.4 Presenting Results of Statistical Analysis

Estimates of lab-to-lab variability and within-lab variability obtained from applying the above statistical modeling procedures are presented within bar charts and tables. Also presented is the percentage that each variance component contributes to total variability (i.e., their sum). Results are presented for each DB testing round, as determined from fitting model (1) in Section 4.4.2 above. In addition, if the likelihood ratio tests described in Section 4.4.3 above determined that the estimates for a specific variance component did not differ significantly across DB testing rounds, the summaries for this component are also presented across the entire pilot study.

5.0 RESULTS

This section reports the results of the proficiency-test sample analyses performed in this double-blind pilot study. Methods discussed in Chapter 4 were used to summarize the data and to conduct statistical modeling to characterize various sources of total variability in the reported data. All references to clients and laboratories in this chapter are made according to the identification protocol introduced in Table 3-1 of Section 3.3.

As discussed in Section 3.3, a total of 12 clients (clients A through L) reported analytical results in this pilot study. These results were obtained from 12 different NLLAP-recognized laboratories: 10 laboratories reporting dust sample results (laboratories 01 through 10), and 11 laboratories reporting paint sample results (laboratories 01 through 09, 11, and 12). Note that nine of the laboratories analyzed both dust and paint samples in this study. Of the 12 laboratories providing analytical results,

- 7 laboratories (01, 03, 05, 06, 07, 09, 10) each received dust samples from one client
- 3 laboratories (02, 04, 08) each received dust samples from two clients
- 9 laboratories (01, 03, 05, 06, 07, 08, 09, 11, 12) each received paint samples from one client
- 2 laboratories (02, 04) each received paint samples from two clients

As mentioned in Section 4.1, this chapter considers the following types of results:

- The measured lead amount in the sample, as reported by the laboratory analyzing the sample and forwarded by the client who was provided the sample (expressed in micrograms of lead for dust samples and percent lead by weight for paint samples).

- Percent recovery associated with the sample, equal to the measured lead amount divided by the target lead amount associated with the bulk source material, as determined in Round 22 of the ELPAT Program (and provided in Table 2-1 of Chapter 2), multiplied by 100%.

5.1 STUDY DATA

5.1.1 Data Corrections

In this pilot study, analytical results associated with the double-blind proficiency-test samples were occasionally reported incorrectly by either the client or the laboratory. When a laboratory found it necessary to correct a result reported earlier to the client but did not identify the error on its own, the original incorrect result was used in the data analyses. This situation occurred with the four dust sample results in DB Round 1 for Client I/Laboratory 07. The laboratory did not consider the dilution factor when originally reporting lead amounts for these samples, and therefore, the results were unusually low relative to the target amount. However, while the laboratory reported corrected values back to the proficiency-testing service at the end of the pilot study after the proficiency-testing service inquired about these results, the original incorrect values were used in the analysis rather than the corrected results. This is because the laboratory would not have corrected these values without the proficiency-testing service's inquiries.

When evidence existed that a client reported results incorrectly relative to the laboratory reports, these results were corrected prior to performing data summaries and analyses. For example, as discussed in Section 3.5, clients occasionally reported the units of analytical results incorrectly on the Sample Tracking and Analysis Report Forms, thereby requiring an adjustment to the affected values. The proficiency-testing service identified such errors upon contacting the clients about unusually high or low results, and the correct results were used in the analysis. However, when evidence did not exist that a given data value was reported incorrectly by the client, the data value was not corrected.

5.1.2 Plots of the Data Values

Figures 5-1 through 5-3 contain plots of measured lead amounts for low-dust, mid-dust, and paint proficiency-test samples, respectively, as reported by the laboratories and their participating clients and after any necessary data corrections were made. Each figure contains three plots, one for each testing round. The horizontal axis of each plot indicates the client/laboratory combination associated with a set of measurements (client ID, followed by lab ID). Each plot includes three horizontal reference lines, with the middle line corresponding to the target lead amount as determined in Round 22 of the ELPAT Program and the other two lines corresponding to the lower and upper limits of the acceptance regions for this ELPAT round. (These numbers were given in Table 2-1 in Chapter 2). The measured lead amounts for proficiency-test samples in the double-blind pilot study are represented by asterisks in these plots, while a laboratory's measured lead amount for the proficiency-test sample it analyzed in Round 22 of the ELPAT Program (a sample produced from the same bulk source material as the double-blind pilot samples) is represented by an open circle. Note that the same ELPAT Round 22 result is portrayed in each set of data corresponding to different clients of the same laboratory.

The vertical axis ranges in Figures 5-1 through 5-3 were selected to allow the variability in results to be seen clearly across the plots. Occasionally, the vertical axis range could not include some very large or very small data values without considerably distorting the plot. Those few results that fell outside of the axis ranges are plotted at the upper (or lower) range limit by an up (or down) arrow and the actual data value(s) in parentheses.

All of the ELPAT Round 22 results plotted in Figures 5-1 through 5-3 were within the acceptance limits determined for that round (i.e., the upper and lower dashed lines in the plots). However, 29 of the 257 double-blind proficiency-test sample results (11%) were outside of these limits. Of these 29 results, 10 were for low-spiked dust-wipe samples from four of the ten

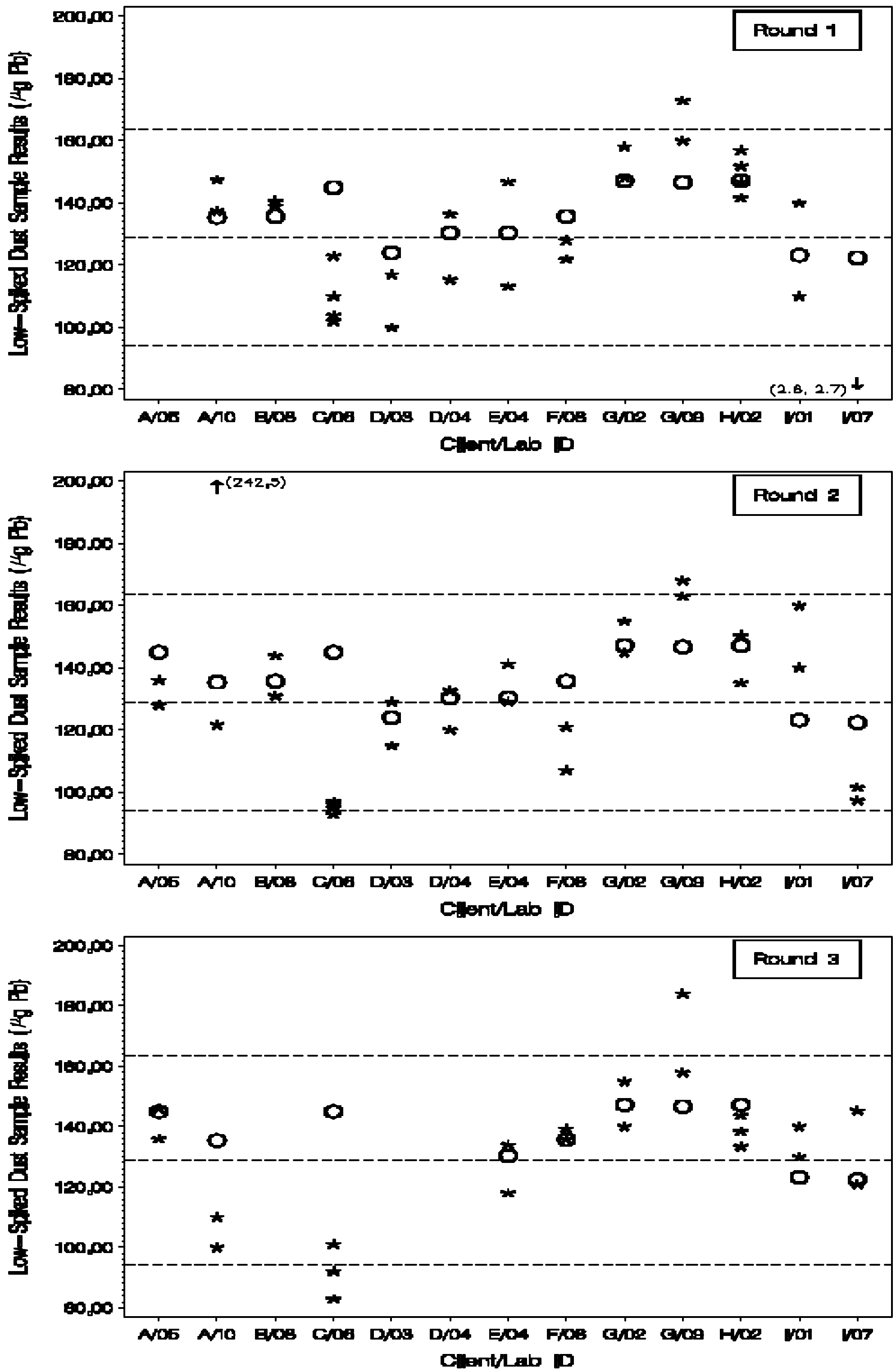

Figure 5-1. Measured Lead Amounts (µg) in the Low-Spiked Dust Proficiency-Test Samples, by Double-Blind Testing Round

(Double-blind results are plotted with asterisks, and results for ELPAT Round 22 are plotted with open circles. Extreme values outside of the vertical axis limits are specified by an arrow and the values in parentheses. The target lead amount (129 µg) and the lower and upper acceptance limits in ELPAT Round 22 are plotted as horizontal dashed lines.)

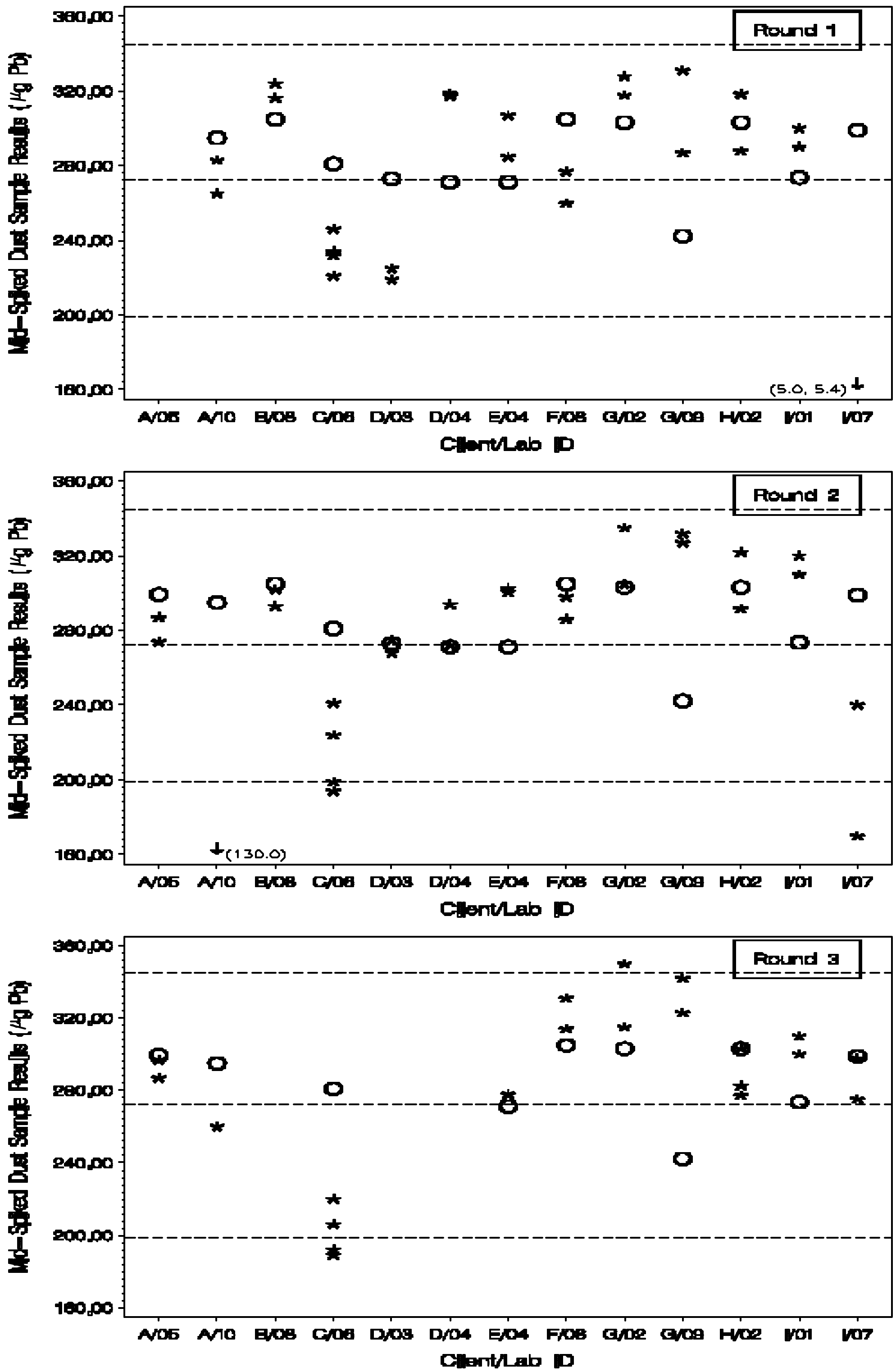

Figure 5-2. Measured Lead Amounts (µg) in the Mid-Spiked Dust Proficiency-Test Samples, by Double-Blind Testing Round

(Double-blind results are plotted with asterisks, and results for ELPAT Round 22 are plotted with open circles. Extreme values outside of the vertical axis limits are specified by an arrow and the values in parentheses. The target lead amount (272 µg) and the lower and upper acceptance limits in ELPAT Round 22 are plotted as horizontal dashed lines.)

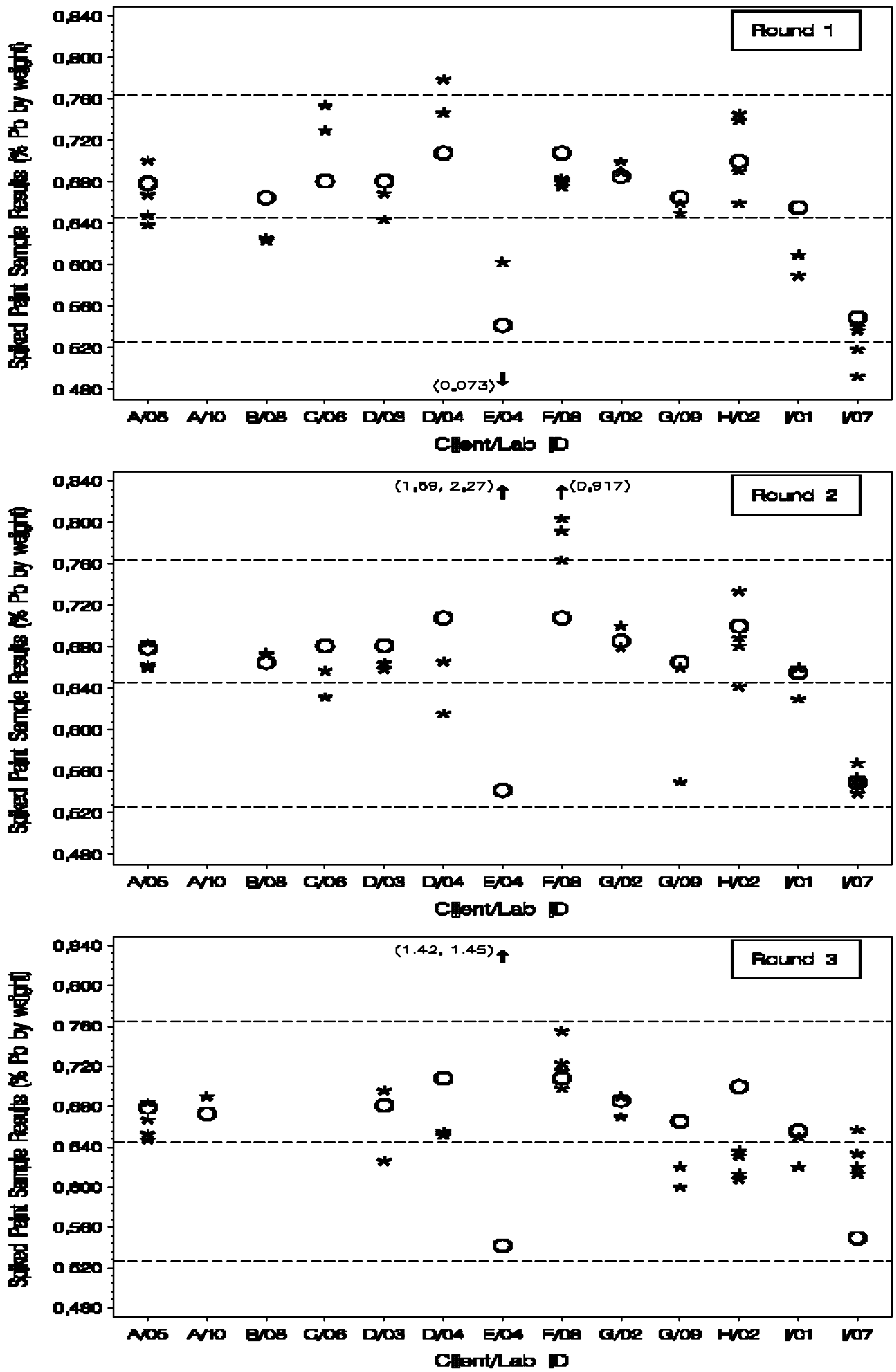

Figure 5-3. Measured Lead Amounts (% by weight) in the Spiked Paint Proficiency-Test Samples, by Double-Blind Testing Round

(Double-blind results are plotted with asterisks, and results for ELPAT Round 22 are plotted with open circles. Extreme values outside of the vertical axis limits are specified by an arrow and the values in parentheses. The target lead amount (0.6454%) and the lower and upper acceptance limits in ELPAT Round 22 are plotted as horizontal dashed lines.)

laboratories analyzing dust-wipe samples, 8 were for mid-spiked dust-wipe samples from four of the ten laboratories analyzing dust-wipe samples, and 11 were for paint chip samples from three of the ten laboratories analyzing paint chip samples. This suggests that for some (but not necessarily all) of the participating laboratories and for each sample type, the double-blind pilot study data are more likely than single-blind study data to exceed the acceptance limits determined from data within the (single-blind) ELPAT Program.

Figures 5-4 through 5-6 contain plots of the percent recoveries (relative to the target lead amount from Round 22 of the ELPAT Program) associated with the low-dust, mid-dust, and paint proficiency-test samples, respectively. The horizontal reference lines in each plot correspond to a percent recovery of 100% and to the ELPAT Round 22 acceptance limits, equal to the acceptance limits plotted in Figures 5-1 through 5-3, respectively, divided by the corresponding target values.

Details on the statistical characterization of the observed levels of variability in these plots and how this variability differs statistically across testing rounds are provided in Sections 5.2 through 5.4.

5.1.3 Identifying Statistical Outliers and Data Exceeding Various Types of Limits

Figures 5-1 through 5-3 show that some data values are very high or very low relative to the other measurements of that sample type (e.g., those values denoted by arrows). Such measurements can greatly influence the results of characterizing variability in the pilot data. To identify those measurements which can be labeled as statistical outliers, the methods of Section 4.2 were applied to all data for each sample type separately. The method could identify up to ten outliers (either high or low) for each sample type.

Table 5-1 contains a list of those data values labeled as statistical outliers by the methods of Section 4.2. These values correspond exactly to those represented by arrows in Figures 5-1 through 5-3 (i.e., were outside of the vertical axis ranges in these plots). The dust sample outliers included the four results from Client I/Laboratory 07 where the dilution factor was not taken into account and two results from Client A/Laboratory 10 whose samples were suspected of having their ID labels switched (although no evidence existed that this actually occurred).

Five of the six paint sample outliers in Table 5-1 were results from Client G/Laboratory 09; these results were as the laboratory reported to the client. Note that all but one of the six paint sample results for Client G/Laboratory 09 were identified as statistical outliers. Client G also sent paint samples to Laboratory 02, but the results for those samples were consistent with other client/laboratory pairs (Figure 5-3). However, as no other clients sent paint samples to Laboratory 09, the extent to which the highly variable results of Laboratory 09 would occur for other clients could not be observed. Note from Figure 5-3 that Laboratory 09 was slightly above the lower limit of the acceptance range for paint samples of the same bulk source material in Round 22 of the ELPAT Program.

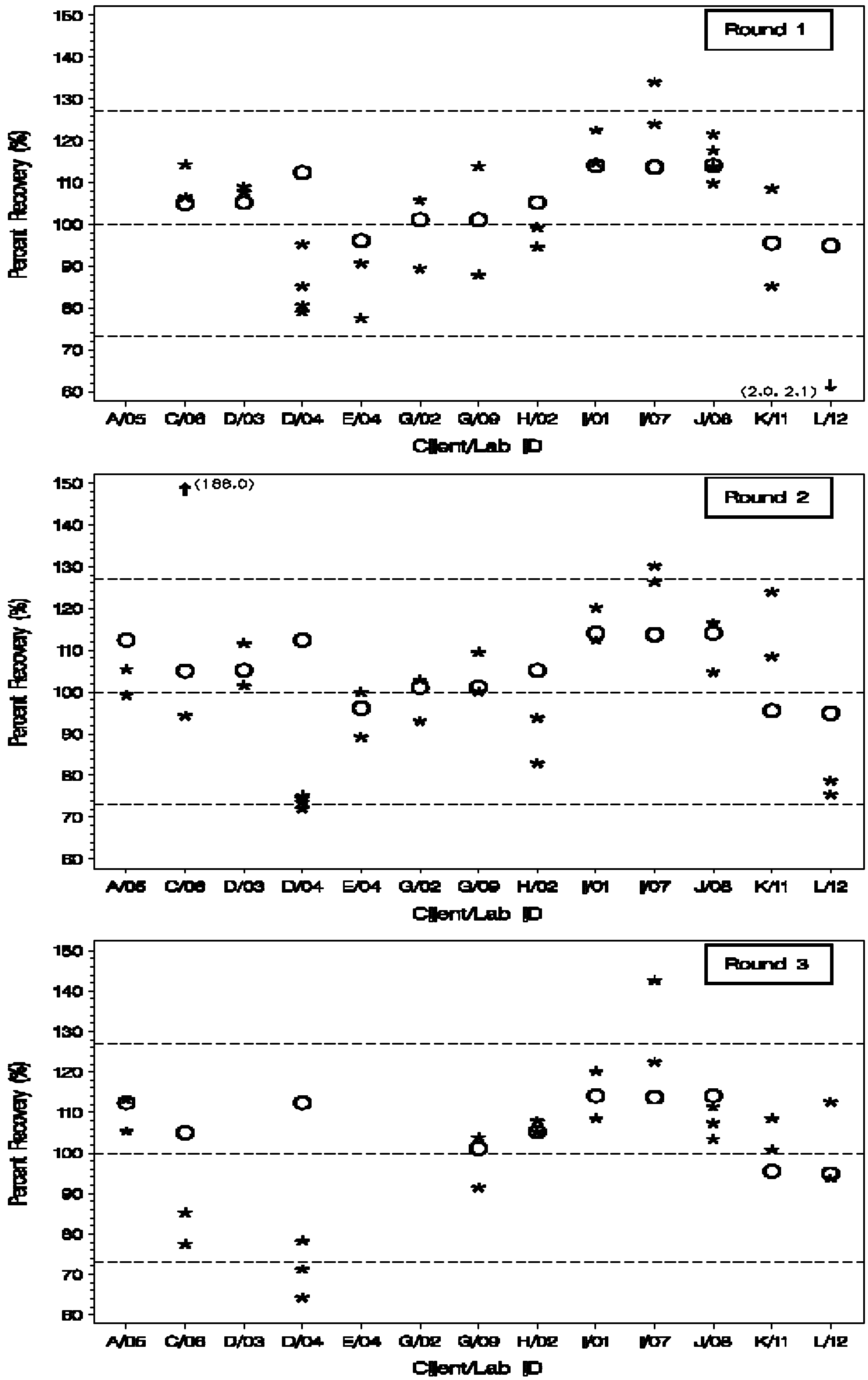

Figure 5-4. **Measured Percent Recoveries in the Low-Spiked Dust Proficiency-Test Samples, by Double-Blind Testing Round**
(Double-blind results are plotted with asterisks, and results for ELPAT Round 22 are plotted with open circles. Extreme values outside of the vertical axis limits are specified by an arrow and the values in parentheses. The percent recovery of 100% and the lower and upper acceptance limits of 73% and 127% in ELPAT Round 22 are plotted as horizontal dashed lines.)

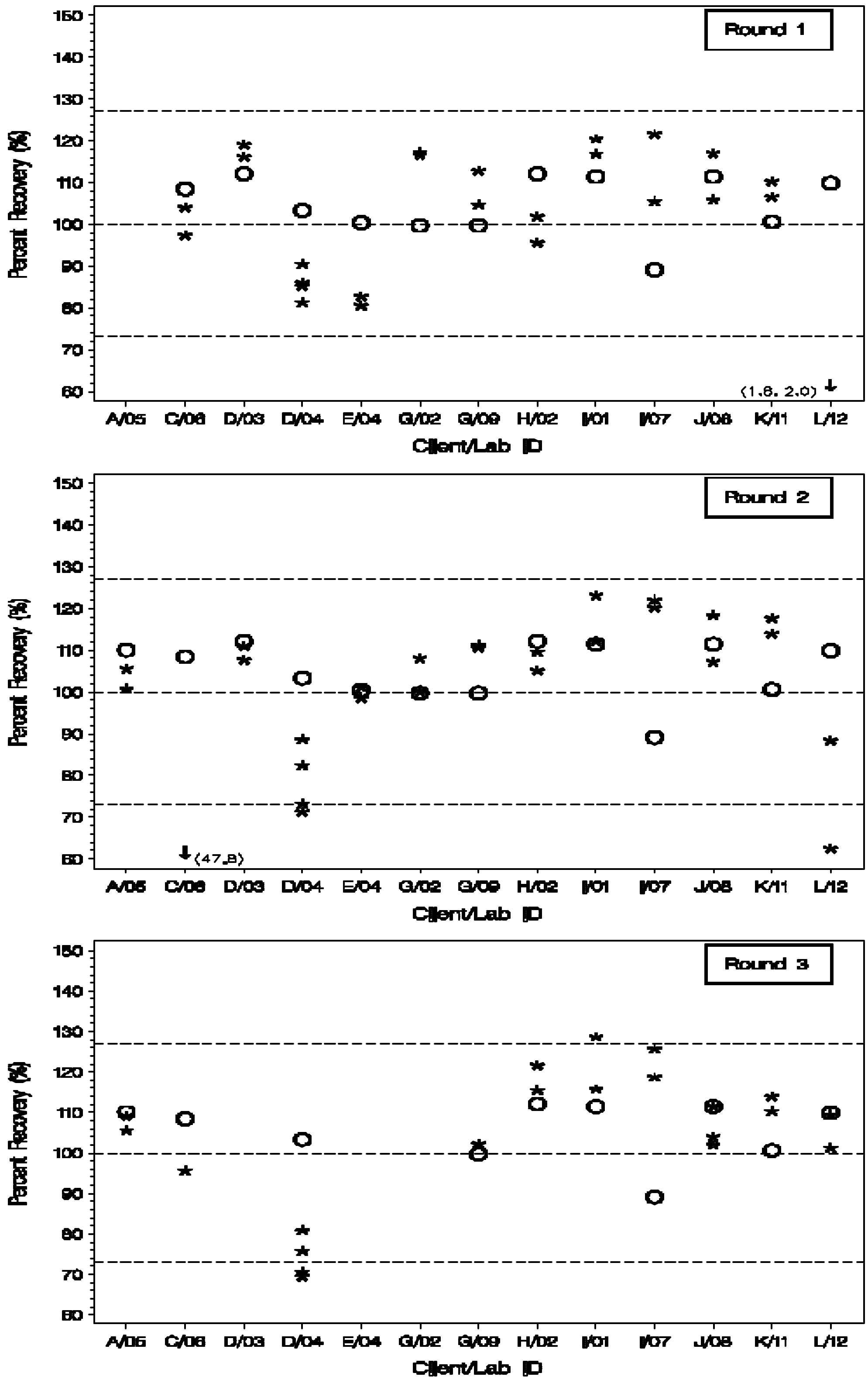

Figure 5-5. Measured Percent Recoveries in the Mid-Spiked Dust Proficiency-Test Samples, by Double-Blind Testing Round

(Double-blind results are plotted with asterisks, and results for ELPAT Round 22 are plotted with open circles. Extreme values outside of the vertical axis limits are specified by an arrow and the values in parentheses. The percent recovery of 100% and the lower and upper acceptance limits of 73% and 127% in ELPAT Round 22 are plotted as horizontal dashed lines.)

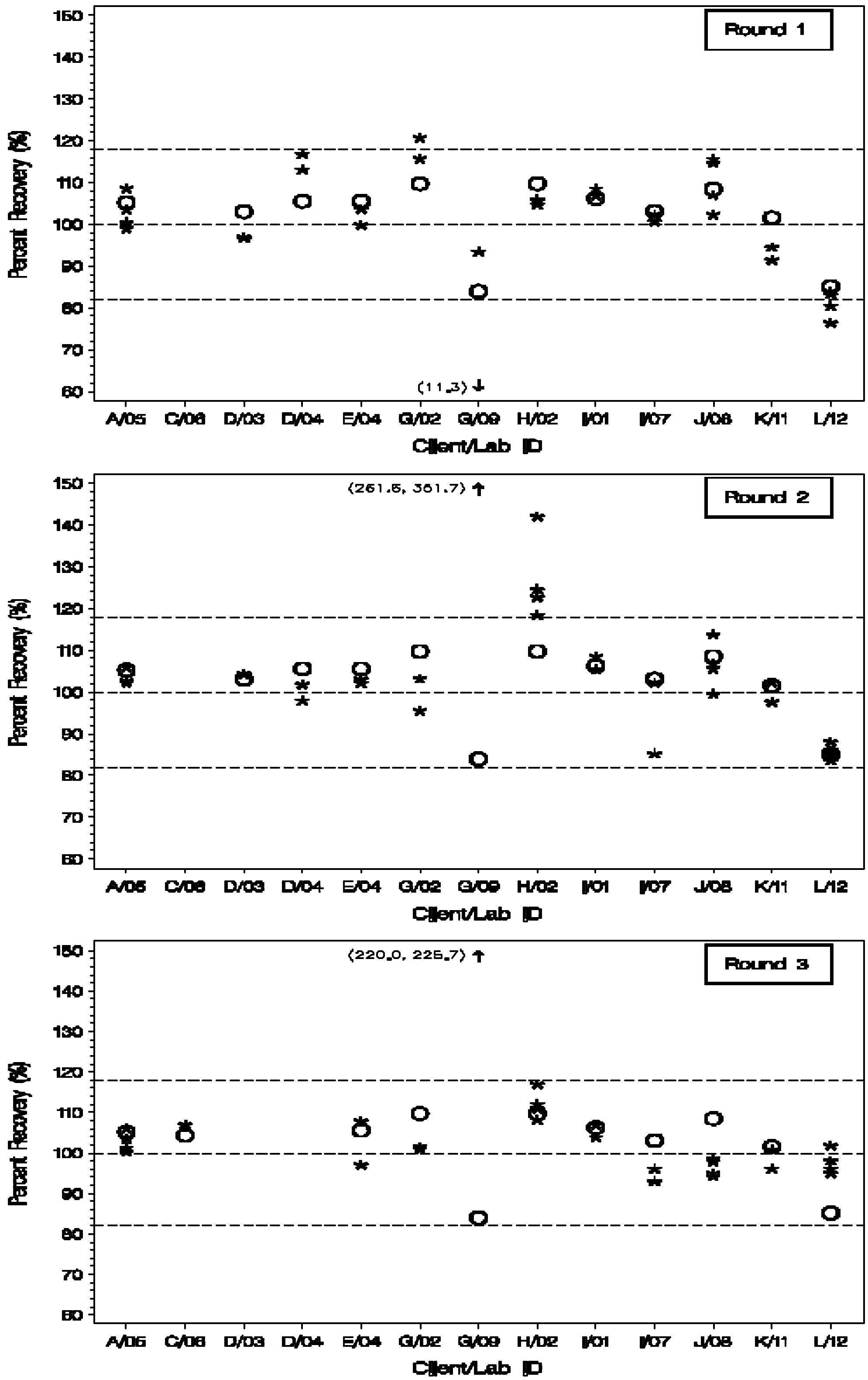

Figure 5-6. Measured Percent Recoveries in the Spiked Paint Proficiency-Test Samples, by Double-Blind Testing Round

(Double-blind results are plotted with asterisks, and results for ELPAT Round 22 are plotted with open circles. Extreme values outside of the vertical axis limits are specified by an arrow and the values in parentheses. The percent recovery of 100% and the lower and upper acceptance limits of 82% and 118% in ELPAT Round 22 are plotted as horizontal dashed lines.)

Table 5-1. Double-Blind Pilot Proficiency-Test Sample Results Identified as Outliers by Applying the Method of Rosner (1983) on All Results by Sample Type[1]

Client/Lab	Testing Round	Reported Lead Amount	Low/High Outlier	% of Target Value[2]
Low-Spiked Dust Samples (n = 82)				
I/07	1	2.60 μg	Low	2.0%
I/07	1	2.70 μg	Low	2.1%
A/10	2	242.5 μg	High	188.0%
Mid-Spiked Dust Samples (n = 82)				
I/07	1	5.00 μg	Low	1.8%
I/07	1	5.40 μg	Low	2.0%
A/10	2	130.0 μg	Low	47.8%
Spiked Paint Samples (n = 93)				
G/09	1	0.073%	Low	11.3%
H/02	2	0.917%	High	142.1%
G/09	3	1.42%	High	220.0%
G/09	3	1.45%	High	224.7%
G/09	2	1.69%	High	261.9%
G/09	2	2.27%	High	351.7%

[1] The method identifies up to 10 outliers for each sample type at a significance level of 0.05 (i.e., the probability of identifying more outliers than are actually present is no higher than 0.05) and assumes a normal distribution.
[2] Target lead amounts are 129 μg for low-dust samples, 272 μg for mid-dust samples, and 0.6454% lead by weight for paint samples.

As a first step to identifying those pilot study data values in Figures 5-1 through 5-6 that would exceed an appropriate acceptance range for a double-blind proficiency-testing program, these data values were compared to a series of ranges that resembled the acceptance limits calculated in Round 22 of the ELPAT Program. These ranges, calculated for each sample type, were as follows:

- the actual acceptance range from Round 22 of the ELPAT Program (documented in Table 2-1 of Chapter 2; the method used to calculate this range is given in Section 2.2)

- a range equal to ±27% (for dust) or ±18.5% (for paint) of the Winsorized mean of the double-blind pilot study data in a given testing round (corresponding to the sizes of the acceptance ranges calculated in Round 22 of the ELPAT Program)

- a range obtained by applying the ELPAT Program method of calculating acceptance ranges on the double-blind pilot study data of a given testing round (i.e., ±3 standard deviations of the mean, calculated after Winsorizing the data)

The latter two ranges were calculated by double-blind testing round and by whether or not the statistical outliers in Table 5-1 were included. Note that these ranges do not necessarily represent appropriate acceptance ranges within a double-blind program, but were considered simply to investigate how the double-blind pilot data may compare to various types of acceptance criteria that are considered in single-blind proficiency testing.

Those double-blind pilot study data values that exceeded at least one of the above ranges are documented in Tables 5-2a through 5-2c for the low-spiked dust, mid-spiked dust, and paint samples, respectively. According to column 4 of Table 5-2a (and as mentioned in Section 5.1.2), 10 of the 82 low-spiked dust sample results exceeded the acceptance range from Round 22 of the ELPAT Program (94-164 µg). Also, 8 of the 82 mid-spiked dust sample results (Table 5-2b) and 11 of the 93 paint sample results (Table 5-2c) exceeded their corresponding acceptance ranges from Round 22 of the ELPAT Program (199-345 µg and 0.5264-0.7645%, respectively). Thus, as discussed in Section 5.1.2, the double-blind pilot study data were more likely than the single-blind data from the ELPAT Program to exceed the acceptance ranges determined from data collected in the (single-blind) ELPAT Program. Note that the inclusion or exclusion of outliers in determining the ranges found in the last two bullets above had only a minor effect on whether non-outliers exceeded these ranges, as the ranges were calculated from Winsorized data. When the methods used to determine acceptance ranges in the ELPAT Program were applied to the double-blind pilot study data, only data values that were among the statistical outliers in Table 5-1 exceeded these ranges.

If the amount of lead spiked into a dust-wipe sample is known, Page 5-33 of the HUD Guidelines (USHUD, 1995) specifies that in a double-blind setting, the laboratory results for these spiked dust-wipe samples should be within 20% of the spiking level. This is more stringent than what was used in ELPAT Round 22 (20% versus 27%). Twenty-seven of 82 low-spiked dust sample results (32.9%) and 17 of 82 mid-spiked dust sample results (20.7%) are more than 20% beyond their respective ELPAT Round 22 target levels. These are considerably more dust sample results than the numbers indicated in the previous paragraph that exceed 27% of the target level (i.e., the ELPAT Round 22 criteria).

Table 5-2a. List of Reported Lead Amounts in Low-Spiked Dust Proficiency-Test Samples That Exceed At Least One of the Various Acceptance Ranges Determined from Methods Used in Round 22 of the ELPAT Program

Client/ Lab	Reported Lead Amount (μg)	Is Result Above or Below the Mean?	Outside of ELPAT Round 22 Acceptance Range?[1]	All Low-Spiked Dust Sample Data Considered		Low-Spiked Dust Sample Data with Outliers Removed (see Table 5-1)	
				Exceeds ±27% of Winsorized Mean of DB Data?[2]	Outside of Acceptance Range as Determined by ELPAT Methods?[3]	Exceeds ±27% of Winsorized Mean of DB Data?[2]	Outside of Acceptance Range as Determined by ELPAT Methods?[3]
Double-Blind Round 1 (n=28)							
I/07	2.60	Below	**T**	**T**	**T**	Outlier removed	
I/07	2.70	Below	**T**	**T**	**T**	Outlier removed	
H/02	156.9	Above		**T**			
G/02	158.1	Above		**T**			
G/09	160.0	Above		**T**			
G/09	173.0	Above	**T**	**T**		**T**	
Double-Blind Round 2 (n=30)							
C/06	93.0	Below	**T**	**T**		**T**	
G/09	168.0	Above	**T**	**T**		**T**	
A/10	242.5	Above	**T**	**T**	**T**	Outlier removed	
Double-Blind Round 3 (n=24)							
C/06	83.0	Below	**T**	**T**		**T**	
C/06	83.0	Below	**T**	**T**		**T**	
C/06	92.0	Below	**T**	**T**		**T**	
G/09	184.0	Above	**T**	**T**		**T**	

[1] Acceptance range in ELPAT Round 22 for these samples was (94 μg, 164 μg), or ±3 standard deviations from the mean (as determined from the Winsorized data of 118 reference laboratories). This acceptance range corresponds to ±27% of the mean.
[2] The observed size of the acceptance range in ELPAT Round 22 for samples generated from this sample batch.
[3] The acceptance range is determined by Winsorizing the lowest and highest 5% of the sample results in the given round, calculating the mean and standard deviation of the Winsorized data, and taking ±3 standard deviations of the mean.

Table 5-2b. List of Reported Lead Amounts in Mid-Spiked Dust Proficiency-Test Samples That Exceed At Least One of the Various Acceptance Ranges Determined from Methods Used in Round 22 of the ELPAT Program

Client/ Lab	Reported Lead Amount (μg)	Is Result Above or Below the Mean?	Outside of ELPAT Round 22 Acceptance Range?[1]	All Mid-Spiked Dust Sample Data Considered		Mid-Spiked Dust Sample Data with Outliers Removed (see Table 5-1)	
				Exceeds ±27% of Winsorized Mean of DB Data?[2]	Outside of Acceptance Range as Determined by ELPAT Methods?[3]	Exceeds ±27% of Winsorized Mean of DB Data?[2]	Outside of Acceptance Range as Determined by ELPAT Methods?[3]
Double-Blind Round 1 (n=28)							
I/07	5.00	Below	**T**	**T**	**T**	Outlier removed	
I/07	5.40	Below	**T**	**T**	**T**	Outlier removed	
Double-Blind Round 2 (n=30)							
A/10	130.0	Below	**T**	**T**	**T**	Outlier removed	
I/07	170.0	Below	**T**	**T**		**T**	
C/06	194.0	Below	**T**	**T**		**T**	
C/06	199.0	Below		**T**		**T**	
Double-Blind Round 3 (n=24)							
C/06	189.0	Below	**T**	**T**		**T**	
C/06	192.0	Below	**T**	**T**		**T**	
G/02	350.0	Above	**T**				

[1] Acceptance range in ELPAT Round 22 for these samples was (199 μg, 345 μg), or ±3 standard deviations from the mean (as determined from the Winsorized data of 118 reference laboratories). This acceptance range corresponds to ±27% of the mean.
[2] The observed size of the acceptance range in ELPAT Round 22 for samples generated from this sample batch.
[3] The acceptance range is determined by Winsorizing the lowest and highest 5% of the sample results in the given round, calculating the mean and standard deviation of the Winsorized data, and taking ±3 standard deviations of the mean.

Table 5-2c. List of Reported Lead Amounts in Spiked Paint Proficiency-Test Samples That Exceed At Least One of the Various Acceptance Ranges Determined from Methods Used in Round 22 of the ELPAT Program

Client/ Lab	Reported Lead Amount (% lead by weight)	Is Result Above or Below the Mean?	Outside of ELPAT Round 22 Acceptance Range?[1]	All Spiked Paint Sample Data Considered		Spiked Paint Sample Data with Outliers Removed (see Table 5-1)	
				Exceeds ±18.5% of Winsorized Mean of DB Data?[2]	Outside of Acceptance Range as Determined by ELPAT Methods?[3]	Exceeds ±18.5% of Winsorized Mean of DB Data?[2]	Outside of Acceptance Range as Determined by ELPAT Methods?[3]
Double-Blind Round 1 (n=32)							
G/09	0.073	Below	**T**	**T**	**T**	Outlier removed	
L/12	0.493	Below	**T**	**T**		**T**	
L/12	0.519	Below	**T**	**T**		**T**	
G/02	0.779	Above	**T**	**T**		**T**	
Double-Blind Round 2 (n=32)							
L/12	0.539	Below		**T**			
L/12	0.547	Below		**T**			
I/07	0.550	Below		**T**			
L/12	0.554	Below		**T**			
H/02	0.792	Above	**T**			**T**	
H/02	0.804	Above	**T**			**T**	
H/02	0.917	Above	**T**	**T**		Outlier removed	
G/09	1.69	Above	**T**	**T**	**T**	Outlier removed	
G/09	2.27	Above	**T**	**T**	**T**	Outlier removed	
Double-Blind Round 3 (n=29)							
G/09	1.42	Above	**T**	**T**	**T**	Outlier removed	
G/09	1.45	Above	**T**	**T**	**T**	Outlier removed	

[1] Acceptance range in ELPAT Round 22 for these samples was (0.5264%, 0.7645%), or ±3 standard deviations from the mean (as determined from the Winsorized data of 118 reference laboratories). This acceptance range corresponds to ±18.5% of the mean.
[2] The observed size of the acceptance range in ELPAT Round 22 for samples generated from this sample batch.
[3] The acceptance range is determined by Winsorizing the lowest and highest 5% of the sample results in the given round, calculating the mean and standard deviation of the Winsorized data, and taking ±3 standard deviations of the mean.

5.2 SUMMARY OF REPORTED LEAD AMOUNTS

This section presents numerical summaries of the data plotted in Figures 5-1 through 5-6. As discussed in Section 4.1, three types of summaries were prepared and presented in tables:

- Individual sample results, summarized across all samples, regardless of client or laboratory
- Individual sample results, summarized by laboratory
- Averages of individual sample results within each client/laboratory combination, summarized across all such combinations.

Summaries across all samples by testing round

Measured lead amounts are summarized across all samples, by double-blind testing round, in Tables 5-3a (low-spiked dust), 5-4a (mid-spiked dust), and 5-5a (paint). The corresponding summaries for percent recovery are found in Tables 5-3b (low-spiked dust), 5-4b (mid-spiked dust), and 5-5b (paint). Summaries are presented both with and without the outliers identified in Table 5-1. Note that the double-blind pilot data summaries in these tables do not consider that some samples were analyzed by the same laboratories but for different clients, while other samples were analyzed by different laboratories.

Also included in Tables 5-3a through 5-5b are data summaries for Round 22 of the ELPAT Program for the same batch of proficiency-test samples, calculated across the following three groups of laboratories: 1) all laboratories participating in ELPAT Round 22, 2) only those laboratories which were NLLAP-recognized (and which acted as "reference laboratories"), and 3) only those laboratories analyzing samples within the double-blind pilot study.

Tables 5-3a through 5-5b show that average percent recoveries for the three sample types and three DB testing rounds ranged from 96.2% to 116.1% when outliers were included, and from 99.9% to 105% when outliers were removed. While excluding the outliers tended to result in mean percent recoveries closer to 100%, this was not the case for mid-dust samples (Table 5-4b), where the percent recoveries in DB Rounds 1 and 2 (in which outliers were identified) averaged higher and slightly farther away from 100% than when the outliers were included. For dust samples, slightly better performance on average was observed in DB Rounds 2 and 3 compared to DB Round 1. Also, when outliers were included, average percent recoveries for paint samples in DB Rounds 2 and 3 were high (Table 5-5b). In general, average percent recoveries for the double-blind testing rounds compared favorably with those reported in ELPAT Round 22.

Table 5-3a. Descriptive Statistics of Reported Lead Amounts (μg) in Low-Spiked Dust Proficiency-Test Samples, Calculated Across Participating Clients and Their Laboratories, by Testing Round

Testing Round	# Samples	Mean	Standard Deviation	Minimum	Percentiles			Maximum
					25th	50th	75th	
All Low-Spiked Dust Sample Data								
DB Round 1	28	124.1	39.4	2.6	111.7	137.0	147.2	173.0
DB Round 2	30	132.7	29.8	93.0	115.0	131.8	145.0	242.5
DB Round 3	24	129.4	24.5	83.0	114.0	135.9	141.9	184.0
Outliers Removed (see Table 5-1)								
DB Round 1	26	133.4	20.1	100.0	115.4	138.3	147.5	173.0
DB Round 2	29	128.9	21.8	93.0	115.0	131.0	144.0	168.0
DB Round 3	24	129.4	24.5	83.0	114.0	135.9	141.9	184.0
ELPAT Round 22 Results for This Sample Batch								
DB labs only[1]	10	135.5	10.1	122.4	124.0	135.6	145.0	147.2
All NLLAP-recognized labs[2]	118	127.9	17.4	2.6	120.1	129.0	138.0	157.0
All ELPAT Round 22 labs	303	125.9	18.1	1.3	118.7	127.0	136.0	180.0

[1] ELPAT Round 22 data for only those laboratories analyzing the given sample type in the double-blind pilot study.
[2] Data are NOT Winsorized.

Table 5-3b. Descriptive Statistics of Reported Percent Recoveries in Low-Spiked Dust Proficiency-Test Samples, Calculated Across Participating Clients and Their Laboratories, by Testing Round[1]

Testing Round	# Samples	Mean	Standard Deviation	Minimum	Percentiles			Maximum
					25th	50th	75th	
All Low-Spiked Dust Sample Data								
DB Round 1	28	96.2	30.5	2.0	86.6	106.2	114.1	134.1
DB Round 2	30	102.9	23.1	72.1	89.1	102.2	112.4	188.0
DB Round 3	24	100.3	19.0	64.3	88.4	105.3	110.0	142.6
Outliers Removed (see Table 5-1)								
DB Round 1	26	103.4	15.6	77.5	89.4	107.2	114.3	134.1
DB Round 2	29	99.9	16.9	72.1	89.1	101.6	111.6	130.2
DB Round 3	24	100.3	19.0	64.3	88.4	105.3	110.0	142.6
ELPAT Round 22 Results for This Sample Batch								
DB labs only[2]	10	105.0	7.9	94.9	96.1	105.1	112.4	114.1
All NLLAP-recognized labs[3]	118	99.1	13.5	2.0	93.1	100.0	107.0	121.7
All ELPAT Round 22 labs	303	97.6	14.0	1.0	92.0	98.4	105.4	139.5

[1] Percent recoveries are calculated relative to the target of 129 μg Pb.
[2] ELPAT Round 22 data for only those laboratories analyzing the given sample type in the double-blind pilot study.
[3] Data are NOT Winsorized.

Table 5-4a. Descriptive Statistics of Reported Lead Amounts (μg) in Mid-Spiked Dust Proficiency-Test Samples, Calculated Across Participating Clients and Their Laboratories, by Testing Round

Testing Round	# Samples	Mean	Standard Deviation	Minimum	Percentiles			Maximum
					25th	50th	75th	
All Mid-Spiked Dust Sample Data								
DB Round 1	28	265.6	81.6	5.0	240.0	287.5	318.1	331.0
DB Round 2	30	276.2	50.6	130.0	250.0	292.4	310.0	335.0
DB Round 3	24	282.1	43.7	189.0	267.7	284.8	312.1	350.0
Outliers Removed (see Table 5-1)								
DB Round 1	26	285.6	36.8	219.0	260.0	289.0	318.3	331.0
DB Round 2	29	281.3	43.1	170.0	268.0	293.0	310.0	335.0
DB Round 3	24	282.1	43.7	189.0	267.7	284.8	312.1	350.0
ELPAT Round 22 Results for This Sample Batch								
DB labs only[1]	10	284.2	19.8	242.3	273.0	288.0	299.3	304.9
All NLLAP-recognized labs[2]	118	294.5	270.2	5.7	257.0	273.8	291.1	3180
All ELPAT Round 22 labs	303	276.4	170.9	2.6	251.0	271.5	288.0	3180

[1] ELPAT Round 22 data for only those laboratories analyzing the given sample type in the double-blind pilot study.
[2] Data are NOT Winsorized.

Table 5-4b. Descriptive Statistics of Reported Percent Recoveries in Mid-Spiked Dust Proficiency-Test Samples, Calculated Across Participating Clients and Their Laboratories, by Testing Round[1]

Testing Round	# Samples	Mean	Standard Deviation	Minimum	Percentiles			Maximum
					25th	50th	75th	
All Mid-Spiked Dust Sample Data								
DB Round 1	28	97.7	30.0	1.8	88.2	105.7	116.9	121.7
DB Round 2	30	101.6	18.6	47.8	91.9	107.5	114.0	123.2
DB Round 3	24	103.7	16.1	69.5	98.4	104.7	114.7	128.7
Outliers Removed (see Table 5-1)								
DB Round 1	26	105.0	13.5	80.5	95.6	106.3	117.0	121.7
DB Round 2	29	103.4	15.9	62.5	98.5	107.7	114.0	123.2
DB Round 3	24	103.7	16.1	69.5	98.4	104.7	114.7	128.7
ELPAT Round 22 Results for This Sample Batch								
DB labs only[2]	10	104.5	7.3	89.1	100.4	105.9	110.0	112.1
All NLLAP-recognized labs[3]	118	108.3	99.3	2.1	94.5	100.7	107.0	1169
All ELPAT Round 22 labs	303	101.6	62.8	1.0	92.3	99.8	105.9	1169

[1] Percent recoveries are calculated relative to the target of 272 μg Pb.
[2] ELPAT Round 22 data for only those laboratories analyzing the given sample type in the double-blind pilot study.
[3] Data are NOT Winsorized.

Table 5-5a. Descriptive Statistics of Reported Lead Amounts (% by weight) in Spiked Paint Proficiency-Test Samples, Calculated Across Participating Clients and Their Laboratories, by Testing Round

Testing Round	# Samples	Mean	Standard Deviation	Minimum	Percentiles			Maximum
					25th	50th	75th	
All Spiked Paint Sample Data								
DB Round 1	32	0.639	0.124	0.073	0.617	0.664	0.696	0.779
DB Round 2	32	0.749	0.340	0.539	0.637	0.664	0.695	2.270
DB Round 3	29	0.710	0.204	0.600	0.626	0.653	0.690	1.450
Outliers Removed (see Table 5-1)								
DB Round 1	31	0.658	0.070	0.493	0.624	0.668	0.700	0.779
DB Round 2	29	0.659	0.066	0.539	0.632	0.661	0.681	0.804
DB Round 3	27	0.656	0.039	0.600	0.620	0.652	0.690	0.755
ELPAT Round 22 Results for This Sample Batch								
DB labs only[1]	11	0.655	0.056	0.542	0.656	0.673	0.686	0.708
All NLLAP-recognized labs[2]	118	0.726	0.881	0.510	0.624	0.651	0.675	10.21
All ELPAT Round 22 labs	323	0.679	0.566	0.063	0.611	0.646	0.677	10.21

[1] ELPAT Round 22 data for only those laboratories analyzing the given sample type in the double-blind pilot study.
[2] Data are NOT Winsorized.

Table 5-5b. Descriptive Statistics of Reported Percent Recoveries in Spiked Paint Proficiency-Test Samples, Calculated Across Participating Clients and Their Laboratories, by Testing Round[1]

Testing Round	# Samples	Mean	Standard Deviation	Minimum	Percentiles			Maximum
					25th	50th	75th	
All Spiked Paint Sample Data								
DB Round 1	32	99.1	19.2	11.3	95.6	102.9	107.8	120.7
DB Round 2	32	116.1	52.7	83.5	98.7	102.9	107.6	351.7
DB Round 3	29	110.0	31.7	93.0	97.0	101.2	106.9	224.7
Outliers Removed (see Table 5-1)								
DB Round 1	31	101.9	10.9	76.4	96.7	103.5	108.5	120.7
DB Round 2	29	102.1	10.2	83.5	97.9	102.4	105.5	124.6
DB Round 3	27	101.7	6.1	93.0	96.1	101.0	106.9	117.0
ELPAT Round 22 Results for This Sample Batch								
DB labs only[2]	11	101.5	8.7	83.9	101.6	104.3	106.3	109.7
All NLLAP-recognized labs[3]	118	112.5	136.6	79.0	96.7	100.8	104.6	1581
All ELPAT Round 22 labs	323	105.2	87.7	9.7	94.6	100.2	104.9	1581

[1] Percent recoveries are calculated relative to the target of 0.6454% lead by weight.
[2] ELPAT Round 22 data for only those laboratories analyzing the given sample type in the double-blind pilot study.
[3] Data are NOT Winsorized.

Laboratory performance in the double-blind testing rounds can also be inferred from Tables 5-3a through 5-5b by observing the variability of the sample results by testing round. Initial review of Figures 5-1 through 5-3 in Section 5.1.2 suggested that for each proficiency-test sample type, data in the double-blind testing rounds had higher variability than for ELPAT Round 22 for the given set of laboratories. For dust-wipe samples, Tables 5-3a and 5-4a confirm this conclusion by noting that even when statistical outliers were excluded, the standard deviation of the dust wipe sample data in each DB testing round was more than twice that reported in ELPAT Round 22 for the same group of laboratories. To compare these variability estimates across the four testing rounds (the three double-blind testing rounds plus ELPAT Round 22), Levene's test for homogeneous variance was applied to these data separately for the low-spiked and mid-spiked dust-wipe samples. While Levene's test indicated that, for both dust-wipe sample types, the differences in these standard deviations across the four testing rounds was not significantly different at the 0.05 level, the standard deviations of the log-transformed dust-wipe data (i.e., the data used in statistical analyses within this report) were significantly different at the 0.05 level across testing rounds. This was true regardless of whether statistical outliers were included when calculating standard deviations within the double-blind testing rounds. The significant difference was primarily due to the lower variability observed in Round 22 of the ELPAT Program relative to the double-blind testing rounds.

Compared to the dust-wipe data, the double-blind proficiency-test paint chip sample data had standard deviations that more closely matched the standard deviation for paint chip sample data from ELPAT Round 22 (Table 5-5a). These standard deviations did not differ significantly (at the 0.05 level) across testing rounds, according to Levene's test.

A more complete statistical evaluation of how variability in proficiency-test sample results differ across testing rounds and of how accurately the laboratories report results near the samples' respective target levels takes into account the different components of variability (lab-to-lab, within-lab) that could be estimated from the double-blind pilot study data. Results of this analysis, and further conclusions that could be made from this analysis on laboratory performance, are presented within Sections 5.3 and 5.4.

Summaries by laboratory for each double-blind testing round

Summaries by laboratory of the double-blind pilot study data are presented according to double-blind testing round in Tables 5-6a (low-spiked dust), 5-7a (mid-spiked dust), and 5-8a (paint) for the measured lead amounts, and in Tables 5-6b (low-spiked dust), 5-7b (mid-spiked dust), and 5-8b (paint) for percent recoveries. Summaries excluding the outliers in Table 5-1 are presented in italics below the summaries that include outliers, for the combinations of laboratories and DB testing rounds that contained outliers. Each laboratory's analysis result for the proficiency-test sample of the same batch that it analyzed in ELPAT Round 22 is included in these tables.

Table 5-6a. Summaries of Reported Lead Amounts (μg) in Low-Spiked Dust Proficiency-Test Samples, by Laboratory and Testing Round

Laboratory ID	ELPAT Round 22 Result	Double-Blind Pilot Study: Mean (Standard Error) (# Samples)[1]		
		Double-Blind Round 1	Double-Blind Round 2	Double-Blind Round 3
01	123.2	125.0 (15.0) (2)	150.0 (10.0) (2)	135.0 (5.0) (2)
02	147.2	150.6 (2.6) (6)	145.2 (3.4) (6)	141.5 (3.0) (6)
03	124.0	108.5 (8.5) (2)	122.0 (7.0) (2)	No data
04	130.4	128.0 (8.2) (4)	130.8 (4.4) (4)	126.0 (7.9) (2)
05	145.0	No data	132.0 (4.0) (2)	141.0 (5.0) (2)
06	145.0	109.8 (4.7) (4)	95.3 (0.9) (4)	89.8 (4.3) (4)
07	122.4	2.7 (0.1) (2) *No data*	99.5 (2.1) (2)	133.1 (12.2) (2)
08	135.7	132.4 (4.5) (4)	125.8 (7.8) (4)	137.6 (1.8) (2)
09	146.7	166.5 (6.5) (2)	165.5 (2.5) (2)	171.0 (13.0) (2)
10	135.4	142.5 (5.0) (2)	182.1 (60.4) (2) *121.7 (--) (1)*	105.0 (5.0) (2)

[1] Results in italics represent summaries with outliers removed (see Table 5-1).

Table 5-6b. Summaries of Reported Percent Recoveries in Low-Spiked Dust Proficiency-Test Samples, by Laboratory and Testing Round[1]

Laboratory ID	ELPAT Round 22 Result	Double-Blind Pilot Study: Mean (Standard Error) (# Samples)[2]		
		Double-Blind Round 1	Double-Blind Round 2	Double-Blind Round 3
01	95.5	96.9 (11.6) (2)	116.3 (7.8) (2)	104.7 (3.9) (2)
02	114.1	116.7 (2.0) (6)	112.6 (2.6) (6)	109.7 (2.3) (6)
03	96.1	84.1 (6.6) (2)	94.6 (5.4) (2)	No data
04	101.1	99.2 (6.3) (4)	101.4 (3.4) (4)	97.7 (6.2) (2)
05	112.4	No data	102.3 (3.1) (2)	109.3 (3.9) (2)
06	112.4	85.1 (3.7) (4)	73.8 (0.7) (4)	69.6 (3.3) (4)
07	94.9	2.1 (0.0) (2) *No data*	77.1 (1.6) (2)	103.2 (9.5) (2)
08	105.2	102.6 (3.5) (4)	97.5 (6.1) (4)	106.6 (1.4) (2)
09	113.7	129.1 (5.0) (2)	128.3 (1.9) (2)	132.6 (10.1) (2)
10	105.0	110.5 (3.9) (2)	141.2 (46.8) (2) *94.3 (--) (1)*	81.4 (3.9) (2)

[1] Percent recoveries are calculated relative to the target of 129 μg Pb.
[2] Results in italics represent summaries with outliers removed (see Table 5-1).

Table 5-7a. Summaries of Reported Lead Amounts (μg) in Mid-Spiked Dust Proficiency-Test Samples, by Laboratory and Testing Round

Laboratory ID	ELPAT Round 22 Result	Double-Blind Pilot Study: Mean (Standard Error) (# Samples)[1]		
		Double-Blind Round 1	Double-Blind Round 2	Double-Blind Round 3
01	273.6	295.0 (5.0) (2)	315.0 (5.0) (2)	305.0 (5.0) (2)
02	303.1	314.7 (5.6) (6)	316.2 (6.3) (6)	301.8 (11.3) (6)
03	273.0	222.0 (3.0) (2)	271.5 (3.5) (2)	No data
04	271.2	307.0 (7.8) (4)	292.3 (7.0) (4)	277.8 (0.2) (2)
05	299.3	No data	280.5 (6.5) (2)	292.0 (5.0) (2)
06	281.0	233.3 (5.1) (4)	214.5 (11.0) (4)	201.8 (7.1) (4)
07	298.9	5.2 (0.2) (2) *No data*	205.2 (35.2) (2)	286.6 (11.2) (2)
08	304.9	294.3 (15.4) (4)	294.8 (3.4) (4)	322.6 (8.3) (2)
09	242.3	309.0 (22.0) (2)	329.5 (2.5) (2)	332.5 (9.5) (2)
10	294.9	274.0 (9.0) (2)	190.0 (60.0) (2) *250.0 (--) (1)*	260.0 (0.0) (2)

[1] Results in italics represent summaries with outliers removed (see Table 5-1).

Table 5-7b. Summaries of Reported Percent Recoveries in Mid-Spiked Dust Proficiency-Test Samples, by Laboratory and Testing Round[1]

Laboratory ID	ELPAT Round 22 Result	Double-Blind Pilot Study: Mean (Standard Error) (# Samples)[2]		
		Double-Blind Round 1	Double-Blind Round 2	Double-Blind Round 3
01	100.6	108.5 (1.8) (2)	115.8 (1.8) (2)	112.1 (1.8) (2)
02	111.4	115.7 (2.0) (6)	116.3 (2.3) (6)	111.0 (4.2) (6)
03	100.4	81.6 (1.1) (2)	99.8 (1.3) (2)	No data
04	99.7	112.9 (2.9) (4)	107.5 (2.6) (4)	102.1 (0.1) (2)
05	110.0	No data	103.1 (2.4) (2)	107.4 (1.8) (2)
06	103.3	85.8 (1.9) (4)	78.9 (4.0) (4)	74.2 (2.6) (4)
07	109.9	1.9 (0.1) (2) *No data*	75.4 (12.9) (2)	105.3 (4.1) (2)
08	112.1	108.2 (5.7) (4)	108.4 (1.3) (4)	118.6 (3.1) (2)
09	89.1	113.6 (8.1) (2)	121.1 (0.9) (2)	122.2 (3.5) (2)
10	108.4	100.7 (3.3) (2)	69.9 (22.1) (2) *91.9 (--) (1)*	95.6 (0.0) (2)

[1] Percent recoveries are calculated relative to the target of 272 μg Pb.
[2] Results in italics represent summaries with outliers removed (see Table 5-1).

Table 5-8a. Summaries of Reported Lead Amounts (% by weight) in Spiked Paint Proficiency-Test Samples, by Laboratory and Testing Round

Laboratory ID	ELPAT Round 22 Result	Double-Blind Pilot Study: Mean (Standard Error) (# Samples)[1]		
		Double-Blind Round 1	Double-Blind Round 2	Double-Blind Round 3
01	0.6858	0.695 (0.005) (2)	0.690 (0.010) (2)	0.680 (0.010) (2)
02	0.7081	0.708 (0.018) (6)	0.760 (0.044) (6) *0.728 (0.037) (5)*	0.700 (0.016) (6)
03	0.6650	0.625 (0.001) (2)	0.674 (0.001) (2)	No data
04	0.6811	0.699 (0.026) (4)	0.653 (0.007) (4)	0.661 (0.035) (2)
05	0.6790	0.664 (0.014) (4)	0.667 (0.006) (4)	0.663 (0.008) (4)
06	0.6730	No data	No data	0.690 (--) (1)
07	0.6653	0.655 (0.005) (2)	0.605 (0.055) (2)	0.610 (0.010) (2)
08	0.7000	0.709 (0.021) (4)	0.687 (0.019) (4)	0.622 (0.007) (4)
09	0.5418	0.338 (0.265) (2) *0.603 (--) (1)*	1.980 (0.290) (2) *No data*	1.435 (0.015) (2) *No data*
11	0.6555	0.600 (0.010) (2)	0.645 (0.015) (2)	0.635 (0.015) (2)
12	0.5493	0.523 (0.011) (4)	0.552 (0.006) (4)	0.631 (0.010) (4)

[1] Results in italics represent summaries with outliers removed (see Table 5-1).

Table 5-8b. Summaries of Reported Percent Recoveries in Spiked Paint Proficiency-Test Samples, by Laboratory and Testing Round[1]

Laboratory ID	ELPAT Round 22 Result	Double-Blind Pilot Study: Mean (Standard Error) (# Samples)[2]		
		Double-Blind Round 1	Double-Blind Round 2	Double-Blind Round 3
01	106.3	107.7 (0.8) (2)	106.9 (1.5) (2)	105.4 (1.5) (2)
02	109.7	109.8 (2.8) (6)	117.7 (6.8) (6) *112.9 (5.8) (5)*	108.4 (2.6) (6)
03	103.0	96.8 (0.2) (2)	104.4 (0.1) (2)	No data
04	105.5	108.4 (4.0) (4)	101.2 (1.1) (4)	102.4 (5.4) (2)
05	105.2	102.9 (2.1) (4)	103.3 (0.9) (4)	102.7 (1.3) (4)
06	104.3	No data	No data	106.9 (--) (1)
07	103.1	101.5 (0.8) (2)	93.7 (8.5) (2)	94.5 (1.5) (2)
08	108.5	109.9 (3.2) (4)	106.4 (2.9) (4)	96.4 (1.1) (4)
09	83.9	52.4 (41.1) (2) *93.4 (--) (1)*	306.8 (44.9) (2) *No data*	222.3 (2.3) (2) *No data*
11	101.6	93.0 (1.5) (2)	99.9 (2.3) (2)	98.4 (2.3) (2)
12	85.1	81.0 (1.7) (4)	85.5 (1.0) (4)	97.7 (1.5) (4)

[1] Percent recoveries are calculated relative to the target of 0.6454% lead by weight.
[2] Results in italics represent summaries with outliers removed (see Table 5-1).

The laboratory summaries (Tables 5-6a through 5-8b) show occasional gaps in the data for certain DB testing rounds, due to reasons such as omitting outliers from analysis (e.g., Laboratory 07 in DB Round 1 for dust) or the client did not report analytical results back for a given set of samples (e.g., Laboratory 03 in DB Round 3 for dust). In general, average lead amounts for a specific laboratory were either consistently above or consistently below the target lead amount across the three DB testing rounds. For example, average percent recoveries for both types of dust samples exceeded 100% in each DB testing round for laboratories 02, 05, and 09, while averages were below 100% in each DB testing round for laboratories 03 and 06 (Tables 5-6b and 5-7b). The unusually high paint results for laboratory 09 relative to the other laboratories are evident throughout the study in Tables 5-8a and 5-8b. Thus, laboratories tended to be consistent in their reported results within and across testing rounds, while results from different laboratories could differ considerably.

<u>Summaries of client/laboratory averages by DB testing round</u>

Summaries of the averages of individual sample results within each client/laboratory combination are presented according to DB testing round in Tables 5-9a (low-spiked dust), 5-10a (mid-spiked dust), and 5-11a (paint) for the measured lead amounts, and Tables 5-9b (low-spiked dust), 5-10b (mid-spiked dust), and 5-11b (paint) for percent recoveries. These tables are similar to Tables 5-3a through 5-5b, but instead of summarizing individual sample results, these results are first averaged within each of the 10-13 client/laboratory combinations, then summarized. Thus, each client/laboratory combination has equal weight, despite one combination being associated with more sample results than another. As in Tables 5-3a through 5-5b, summaries are presented both with and without the outliers in Table 5-1 included.

While the reported means in these tables are similar to those in Tables 5-1 and 5-2, the variability in the results is slightly lower due to summarizing average lead measurements rather than measurements for individual samples. (Averages have lower variability than the individual data values used to calculate them, as more information is used to determine the average.) The summaries show that averages tended to be higher than their respective target amounts in each DB testing round when outliers were removed, with similar performance noted from one DB testing round to the next by the laboratories.

5.3 CHARACTERIZING COMPONENTS OF VARIATION

The tables and figures in Sections 5.1 and 5.2 provided preliminary information on how results in the double-blind pilot study differed from one laboratory to another and from one testing round to another. In addition, the figures provided some initial information on the two sources of variability in the data within a testing round: lab-to-lab variability and within-lab variability. This section uses statistical modeling techniques documented in Section 4.4 to characterize lab-to-lab and within-lab variability in the double-blind pilot study data and to determine how this variability differs across the three double-blind testing rounds for a given proficiency-test sample type.

Table 5-9a. Descriptive Statistics of Average Reported Lead Amounts (μg) in Low-Spiked Dust Proficiency-Test Samples for Each Client/ Laboratory Combination, by Double-Blind Testing Round

Testing Round	# Client/Lab Combinations	Average Measured Lead Amount in Low-Spiked Dust Samples for Each Client/Laboratory Combination (μg)						
		Mean	Standard Deviation	Minimum	Percentiles			Maximum
					25th	50th	75th	
All Low-Spiked Dust Sample Data								
DB Round 1	12	123.2	41.7	2.7	117.4	128.0	145.9	166.5
DB Round 2	13	134.8	24.5	95.3	122.0	135.2	150.0	182.1
DB Round 3	10	132.4	22.3	89.8	126.0	136.3	141.0	171.0
Outliers Removed (see Table 5-1)								
DB Round 1	11	134.1	18.0	108.5	125.0	130.1	149.3	166.5
DB Round 2	13	130.1	20.1	95.3	121.7	132.0	142.8	165.5
DB Round 3	10	132.4	22.3	89.8	126.0	136.3	141.0	171.0

Table 5-9b. Descriptive Statistics of Average Reported Percent Recovery in Low-Spiked Dust Proficiency-Test Samples for Each Client/Laboratory Combination, by Double-Blind Testing Round[1]

Testing Round	# Client/Lab Combinations	Average Percent Recovery for Low-Spiked Dust Samples for Each Client/Laboratory Combination						
		Mean	Standard Deviation	Minimum	Percentiles			Maximum
					25th	50th	75th	
All Low-Spiked Dust Sample Data								
DB Round 1	12	95.5	32.3	2.1	91.0	99.2	113.1	129.1
DB Round 2	13	104.5	19.0	73.8	94.6	104.8	116.3	141.2
DB Round 3	10	102.7	17.3	69.6	97.7	105.6	109.3	132.6
Outliers Removed (see Table 5-1)								
DB Round 1	11	104.0	14.0	84.1	96.9	100.9	115.7	129.1
DB Round 2	13	100.9	15.6	73.8	94.3	102.3	110.7	128.3
DB Round 3	10	102.7	17.3	69.6	97.7	105.6	109.3	132.6

[1] Percent recoveries are calculated relative to the target of 129 μg Pb.

Table 5-10a. Descriptive Statistics of Average Reported Lead Amounts (μg) in Mid-Spiked Dust Proficiency-Test Samples for Each Client/ Laboratory Combination, by Double-Blind Testing Round

Testing Round	# Client/Lab Combinations	Average Measured Lead Amount in Mid-Spiked Dust Samples for Each Client/Laboratory Combination (μg)						
		Mean	Standard Deviation	Minimum	Percentiles			Maximum
					25th	50th	75th	
All Mid-Spiked Dust Sample Data								
DB Round 1	12	264.5	88.2	5.2	250.9	295.5	314.4	322.8
DB Round 2	13	278.0	46.0	190.0	271.5	292.0	314.4	329.5
DB Round 3	10	289.7	39.1	201.8	277.8	289.3	322.6	332.5
Outliers Removed (see Table 5-1)								
DB Round 1	11	288.1	34.8	222.0	268.5	296.0	318.1	322.8
DB Round 2	13	282.7	38.9	205.2	271.5	292.0	314.4	329.5
DB Round 3	10	289.7	39.1	201.8	277.8	289.3	322.6	332.5

Table 5-10b. Descriptive Statistics of Average Reported Percent Recovery in Mid-Spiked Dust Proficiency-Test Samples for Each Client/ Laboratory Combination, by Double-Blind Testing Round[1]

Testing Round	# Client/Lab Combinations	Average Percent Recovery for Mid-Spiked Dust Samples for Each Client/Laboratory Combination						
		Mean	Standard Deviation	Minimum	Percentiles			Maximum
					25th	50th	75th	
All Mid-Spiked Dust Sample Data								
DB Round 1	12	97.3	32.4	1.9	92.2	108.6	115.6	118.7
DB Round 2	13	102.2	16.9	69.9	99.8	107.4	115.6	121.1
DB Round 3	10	106.5	14.4	74.2	102.1	106.4	118.6	122.2
Outliers Removed (see Table 5-1)								
DB Round 1	11	105.9	12.8	81.6	98.7	108.8	116.9	118.7
DB Round 2	13	103.9	14.3	75.4	99.8	107.4	115.6	121.1
DB Round 3	10	106.5	14.4	74.2	102.1	106.4	118.6	122.2

[1] Percent recoveries are calculated relative to the target of 272 μg Pb.

Table 5-11a. Descriptive Statistics of Average Reported Lead Amounts (% by weight) in Spiked Paint Proficiency-Test Samples for Each Client/Laboratory Combination, by Double-Blind Testing Round

Testing Round	# Client/Lab Combinations	Average Measured Lead Amount in Spiked Paint Samples for Each Client/Laboratory Combination (% by weight)						
		Mean	Standard Deviation	Minimum	Percentiles			Maximum
					25th	50th	75th	
All Spiked Paint Sample Data								
DB Round 1	12	0.638	0.114	0.338	0.613	0.660	0.702	0.763
DB Round 2	12	0.772	0.385	0.552	0.643	0.665	0.688	1.980
DB Round 3	11	0.728	0.237	0.610	0.631	0.661	0.690	1.435
Outliers Removed (see Table 5-1)								
DB Round 1	12	0.660	0.066	0.523	0.614	0.660	0.702	0.763
DB Round 2	11	0.659	0.058	0.552	0.641	0.662	0.687	0.787
DB Round 3	10	0.657	0.034	0.610	0.631	0.657	0.680	0.723

Table 5-11b. Descriptive Statistics of Average Reported Percent Recovery in Spiked Paint Proficiency-Test Samples for Each Client/Laboratory Combination, by Double-Blind Testing Round[1]

Testing Round	# Client/Lab Combinations	Average Percent Recovery for Spiked Paint Samples for Each Client/Laboratory Combination						
		Mean	Standard Deviation	Minimum	Percentiles			Maximum
					25th	50th	75th	
All Spiked Paint Sample Data								
DB Round 1	12	98.8	17.7	52.4	94.9	102.3	108.8	118.2
DB Round 2	12	119.6	59.7	85.5	99.6	103.0	106.6	306.8
DB Round 3	11	112.7	36.7	94.5	97.7	102.4	106.9	222.3
Outliers Removed (see Table 5-1)								
DB Round 1	12	102.2	10.3	81.0	95.1	102.3	108.8	118.2
DB Round 2	11	102.2	8.9	85.5	99.3	102.6	106.4	121.9
DB Round 3	10	101.8	5.3	94.5	97.7	101.8	105.4	112.0

[1] Percent recoveries are calculated relative to the target of 0.6454% lead by weight.

To portray the observed distributions of the entire pilot study data graphically, Figures 5-7 through 5-9 contain histograms of these data, with data for the entire study (i.e., all three testing rounds) summarized within a single histogram for each sample type. The double-blind pilot data are represented in these histograms within unshaded bars. To compare variability in the results between the double-blind pilot and Round 22 of the ELPAT Program for the participating laboratories, the histograms include the ELPAT Round 22 data within shaded bars.

Note from Figures 5-7 through 5-9 that although the double-blind pilot study has considerably more data than Round 22 of the ELPAT Program for the participating laboratories, total variability in the double-blind pilot data is considerably greater than in ELPAT Round 22 for each sample type. While this is especially true for the dust sample results, the outliers identified in Table 5-1 are clearly visible within the paint histogram and result in most of the variability observed in that histogram. These figures suggest that additional variability is present in the double-blind pilot study data compared to the single-blind data for the group of laboratories participating in this pilot study. However, as seen in Tables 5-3a through 5-5a when considering the larger group of 118 NLLAP-recognized laboratories, it is possible to observe extreme and wide-ranging data values even within a testing round of the single-blind ELPAT Program.

Statistical modeling results

Section 4.4 presented the random effects modeling approach taken to characterize lab-to-lab and within-lab variability across testing rounds in the reported lead amounts for the proficiency-test samples analyzed in this double-blind pilot study. Model (1) in Section 4.4 was fitted to these data for each of the three proficiency-test sample type, both with and without the outliers in Table 5-1 included, for a total of six fits. Recall that this model assumed that each sample result was equal to an overall predicted value, which could change from one testing round to the next, plus or minus some random amount that comes from a normal distribution with mean zero and positive variance. This variance was assumed to contain lab-to-lab and within-lab components which were each allowed to change from one testing round to the next. To test whether the estimates of lab-to-lab variability and within-lab variability differed from one double-blind testing round to the next, reduced versions of model (1) were fitted to the same set of data, and the estimated likelihoods were statistically compared among model (1) and the reduced models. Dust-lead measurements were logarithmically-transformed prior to performing this analysis.

The results of fitting model (1) to the double-blind pilot study data are presented graphically in Figures 5-10 through 5-15 and are documented in Table 5-12. Figures 5-10 through 5-15 each contain two sets of bar charts, one representing an analysis performed without the statistical outliers in Table 5-1 included, and the other representing an analysis performed on all data. For each DB testing round, Figures 5-10 through 5-12 plot the percentages of total variability attributable to each of its two assumed components, lab-to-lab variability (solid portion of each bar) and within-lab variability ("cross-hatched" portion of each bar), for low-

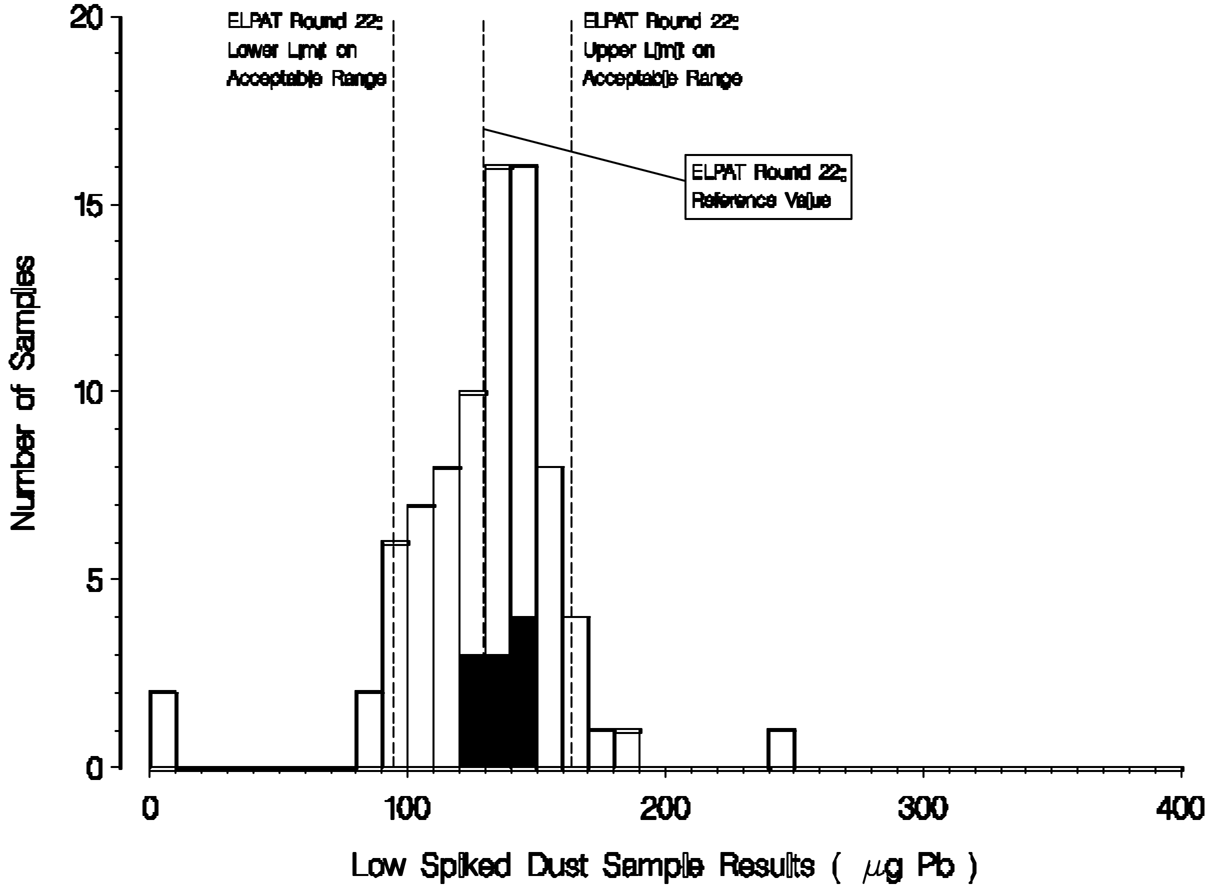

Figure 5-7. **Histogram of Low-Spiked Dust Sample Results (µg) Reported in the Double-Blind Pilot Study (unshaded bars) and in Round 22 of the ELPAT Program (shaded bars) by Laboratories Participating in the Pilot**

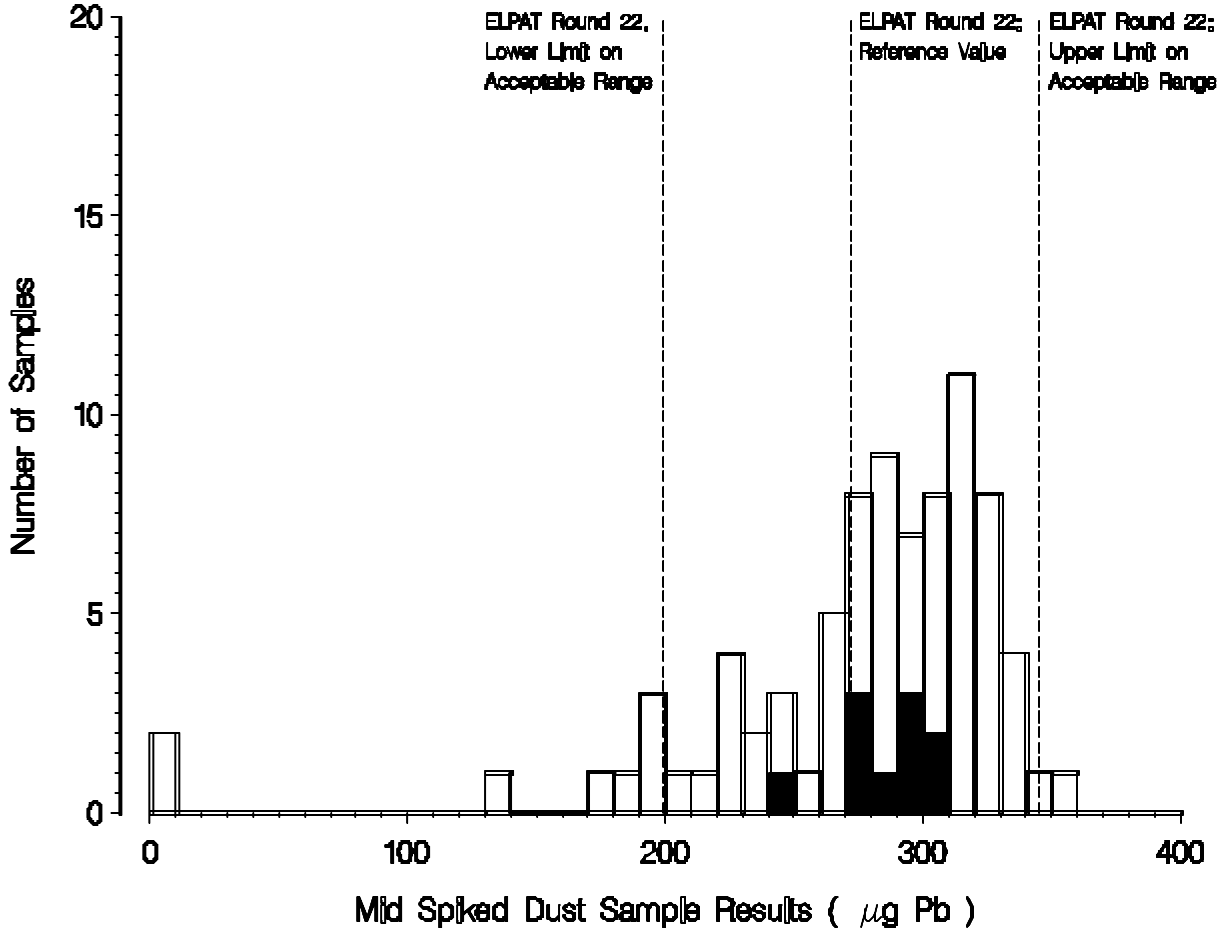

Figure 5-8. Histogram of Mid-Spiked Dust Sample Results (µg) Reported in the Double-Blind Pilot Study (unshaded bars) and in Round 22 of the ELPAT Program (shaded bars) by Laboratories Participating in the Pilot

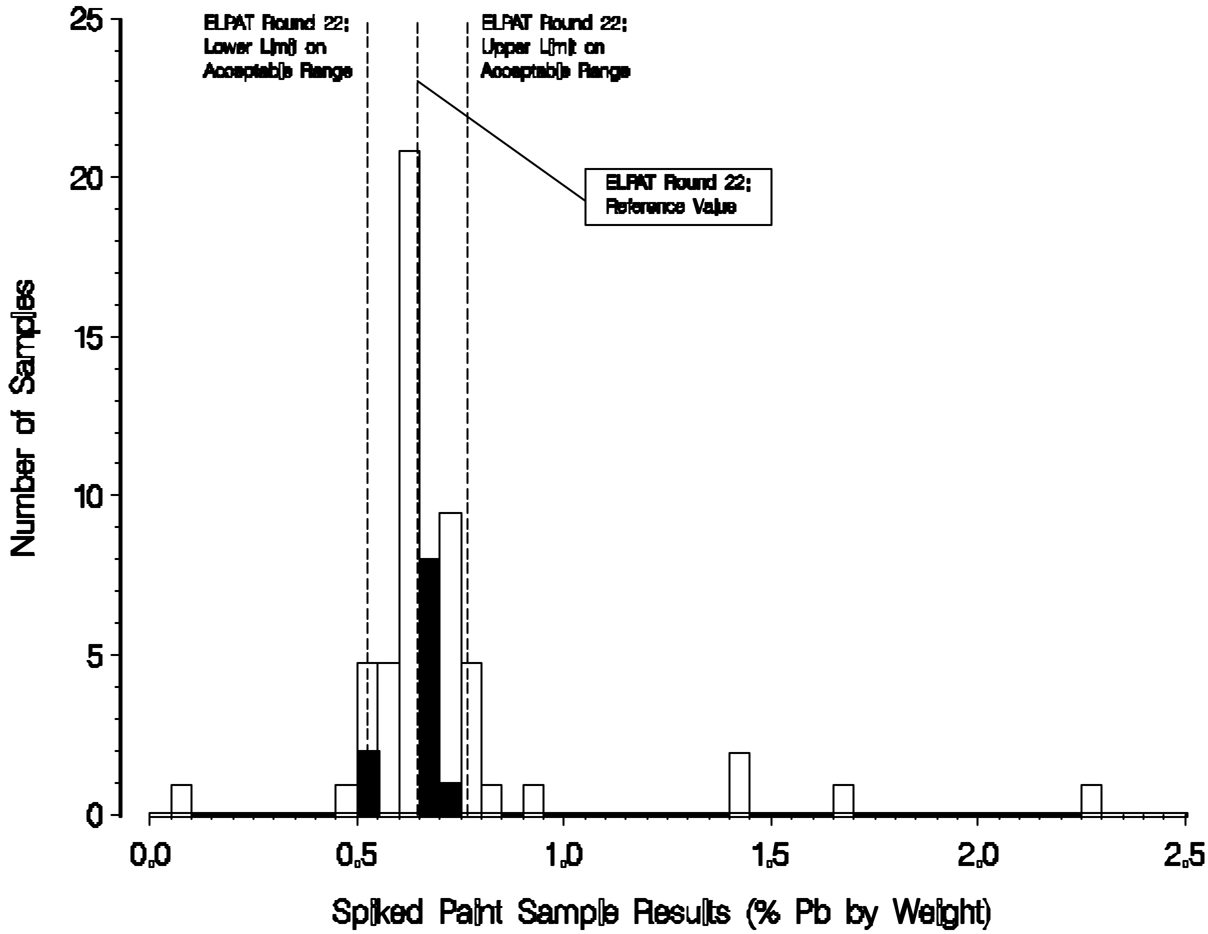

Figure 5-9. Histogram of Spiked Paint Sample Results (% by weight) Reported in the Double-Blind Pilot Study (unshaded bars) and in Round 22 of the ELPAT Program (shaded bars) by Laboratories Participating in the Pilot

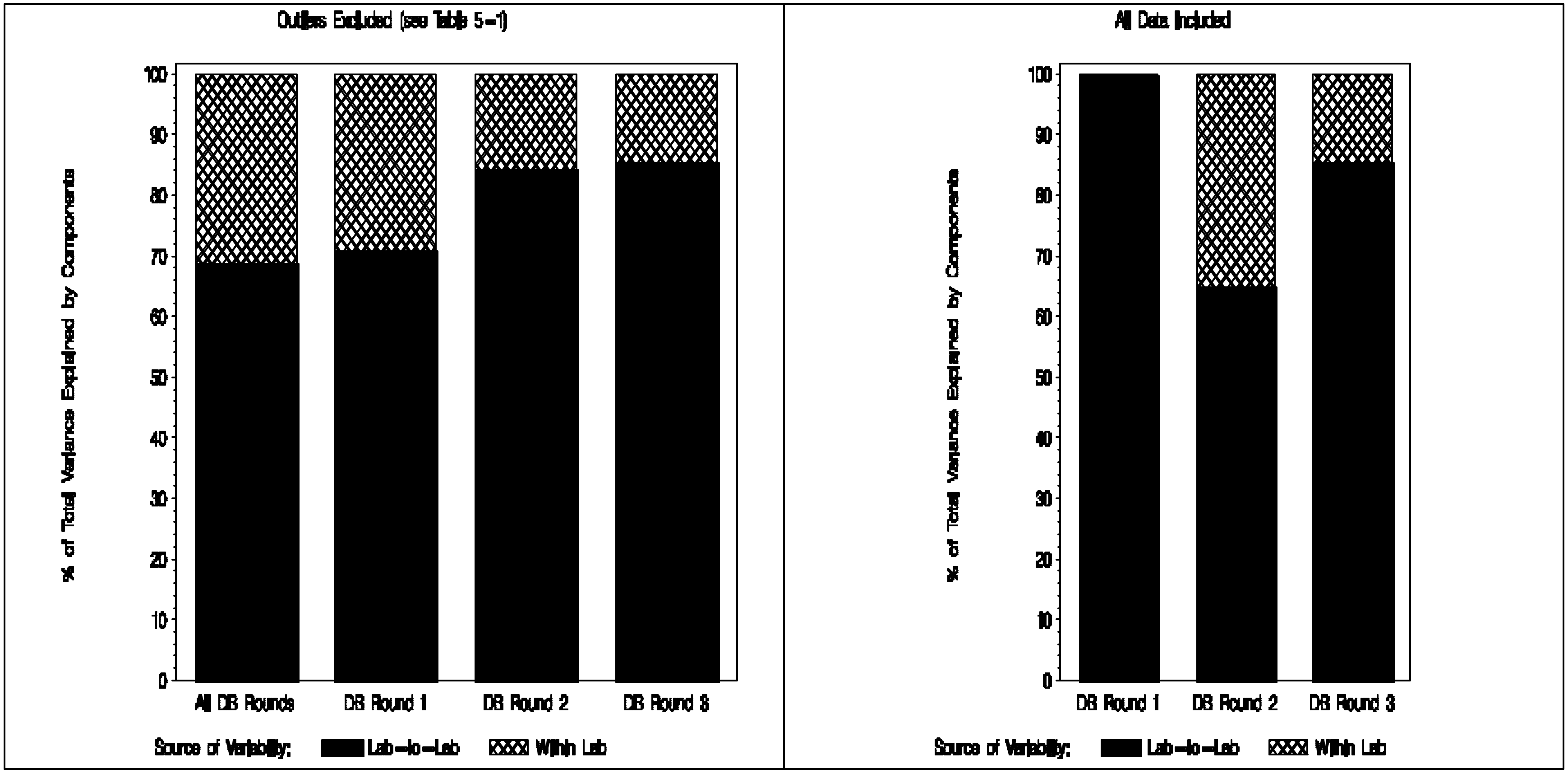

Figure 5-10. Estimated Percentage of Total Variability in Reported Lead Amounts Within Low-Spiked Dust Samples That is Attributable to Lab-to-Lab and Within-Lab Sources, Based on Analyses Performed With and Without Statistical Outliers Included

(Note: "All DB Rounds" represents estimated percentages over the entire pilot study. These estimates were calculated only when statistical outliers were removed, as the percentages differed significantly from one DB testing round to another when the outliers were included.)

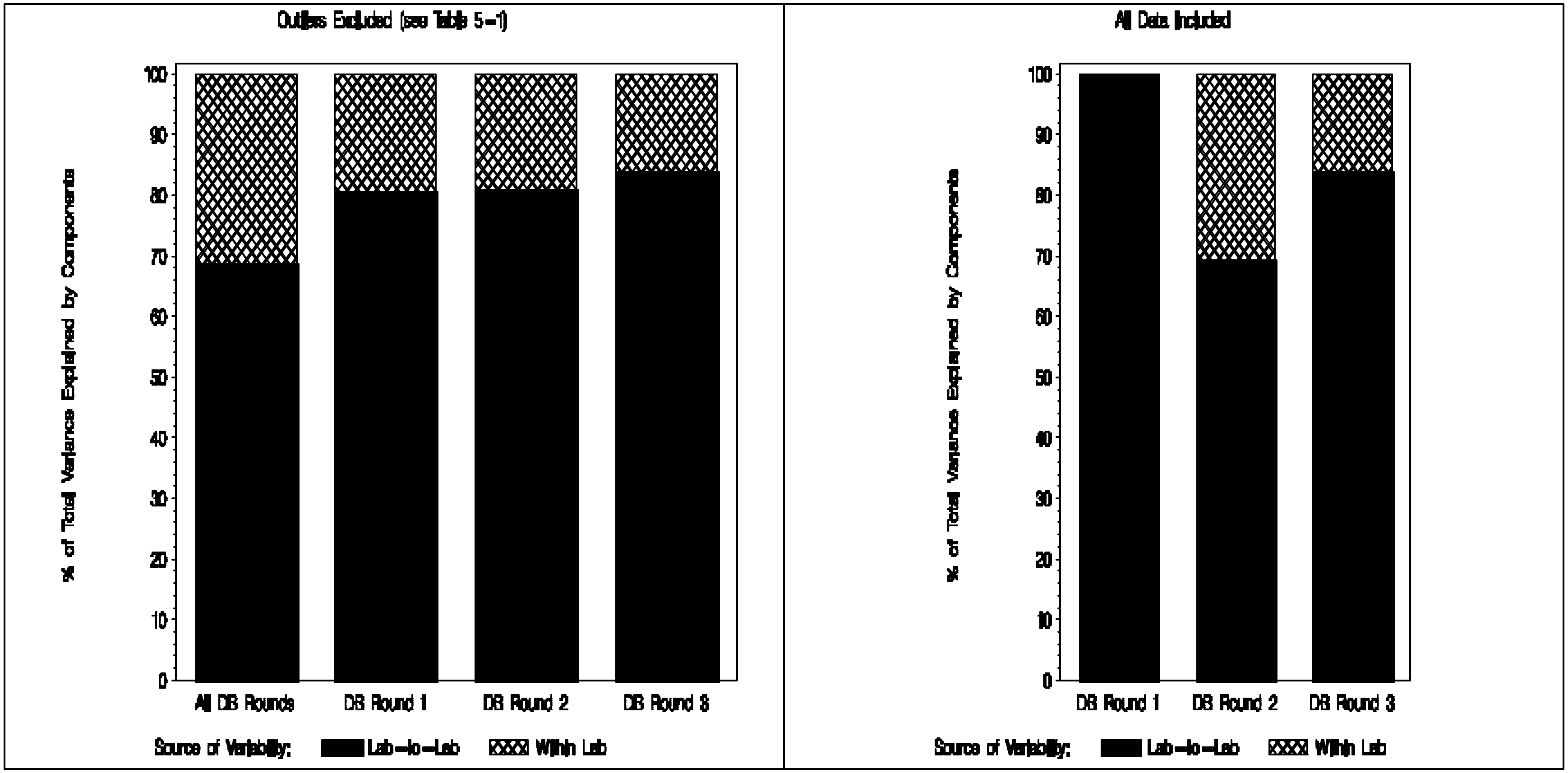

Figure 5-11. Estimated Percentage of Total Variability in Reported Lead Amounts Within Mid-Spiked Dust Samples That is Attributable to Lab-to-Lab and Within-Lab Sources, Based on Analyses Performed With and Without Statistical Outliers Included

(Note: "All DB Rounds" represents estimated percentages over the entire pilot study. These estimates were calculated only when statistical outliers were removed, as the percentages differed significantly from one DB testing round to another when the outliers were included.)

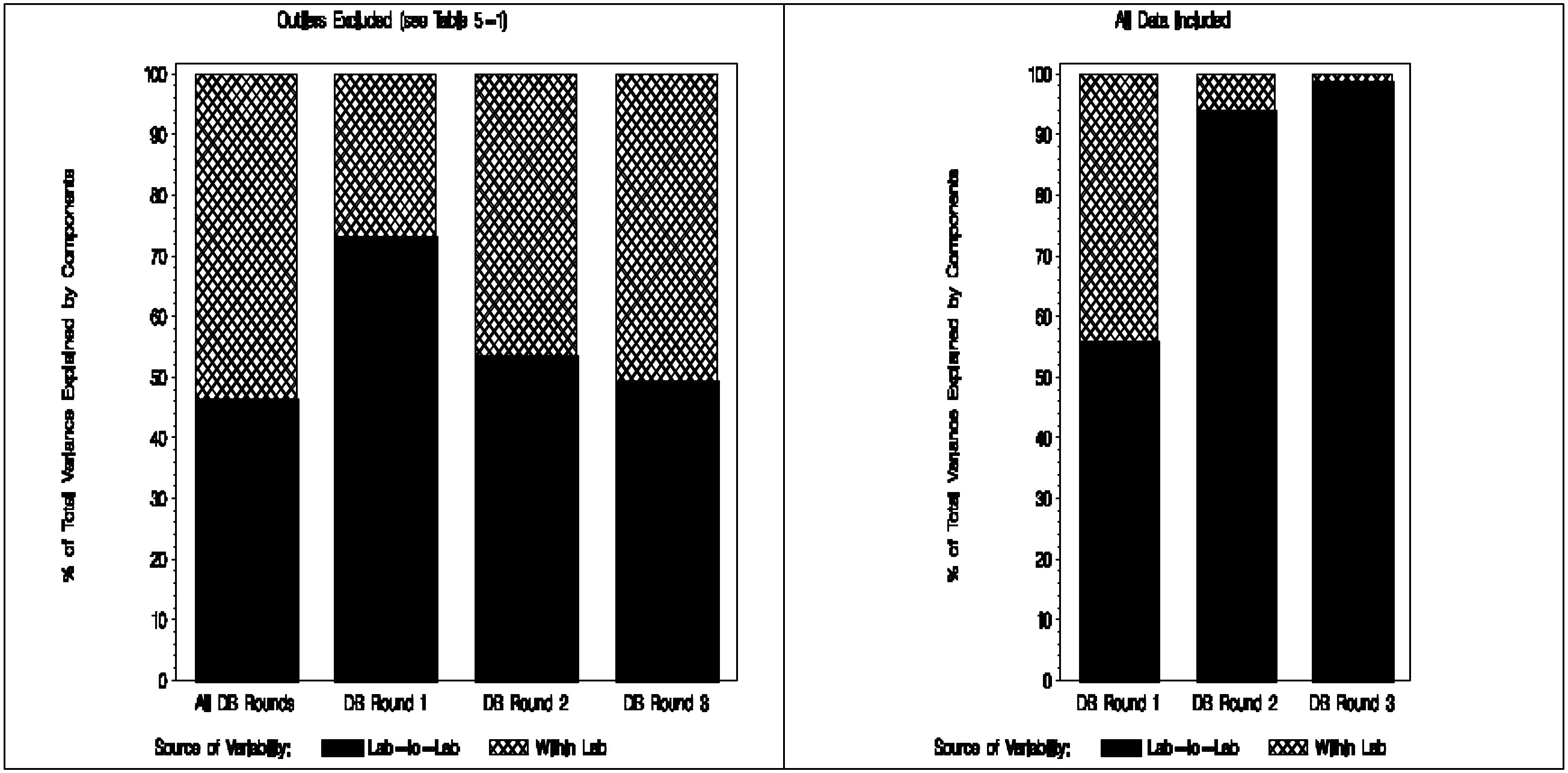

Figure 5-12. Estimated Percentage of Total Variability in Reported Lead Amounts Within Spiked Paint Samples That is Attributable to Lab-to-Lab and Within-Lab Sources, Based on Analyses Performed With and Without Statistical Outliers Included

(Note: "All DB Rounds" represents estimated percentages over the entire pilot study. These estimates were calculated only when statistical outliers were removed, as the percentages differed significantly from one DB testing round to another when the outliers were included.)

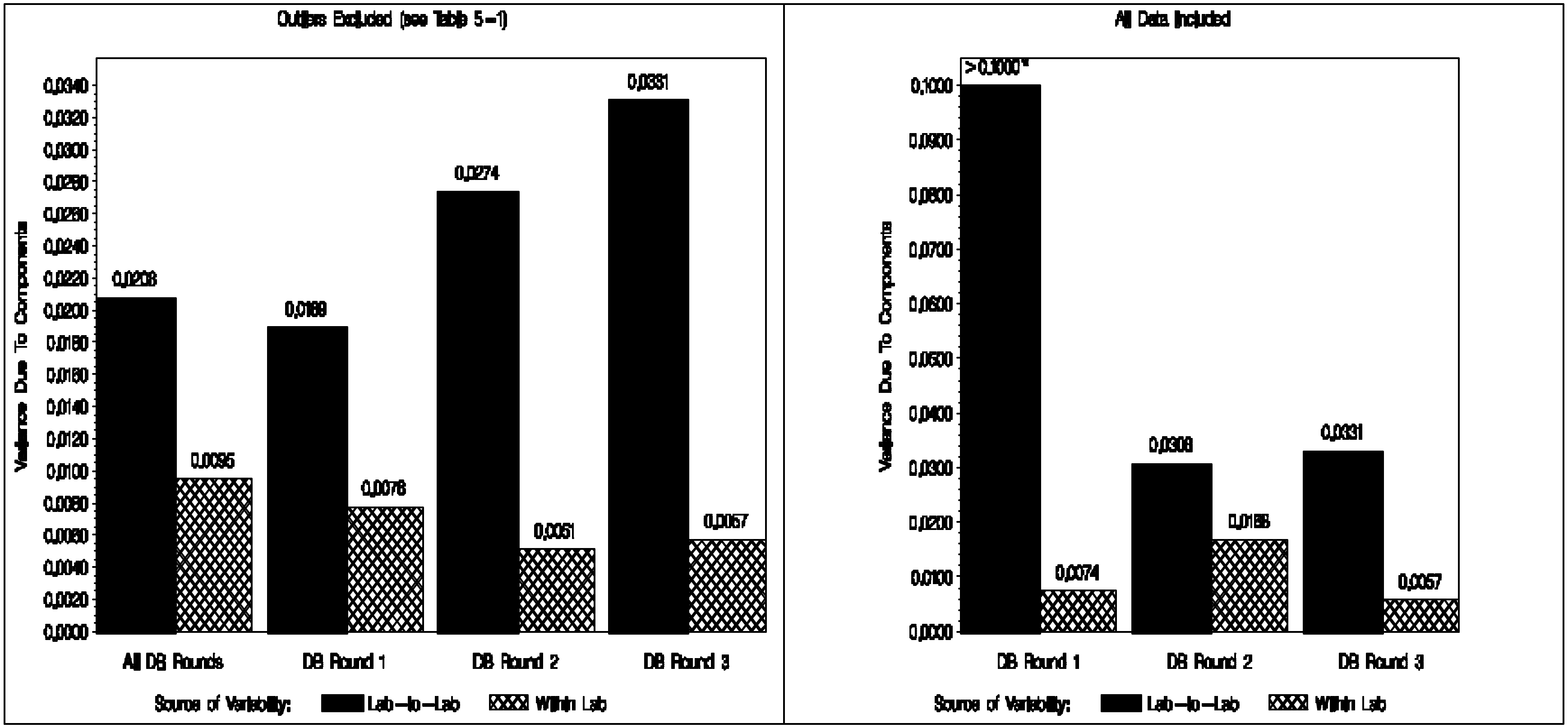

Figure 5-13. Estimates of Lab-to-Lab and Within-Lab Sources of Variability in Reported Lead Amounts Within Low-Spiked Dust Samples, Based on Analyses Performed With and Without Statistical Outliers Included

Note: Vertical axis is in $(\log(\mu g))^2$.
Note: "All DB Rounds" represents estimates over the entire pilot study. These estimates were calculated only when statistical outliers were removed, as the estimates differed significantly from one DB testing round to another when the outliers were included.

* Value of estimate is 1.7075.

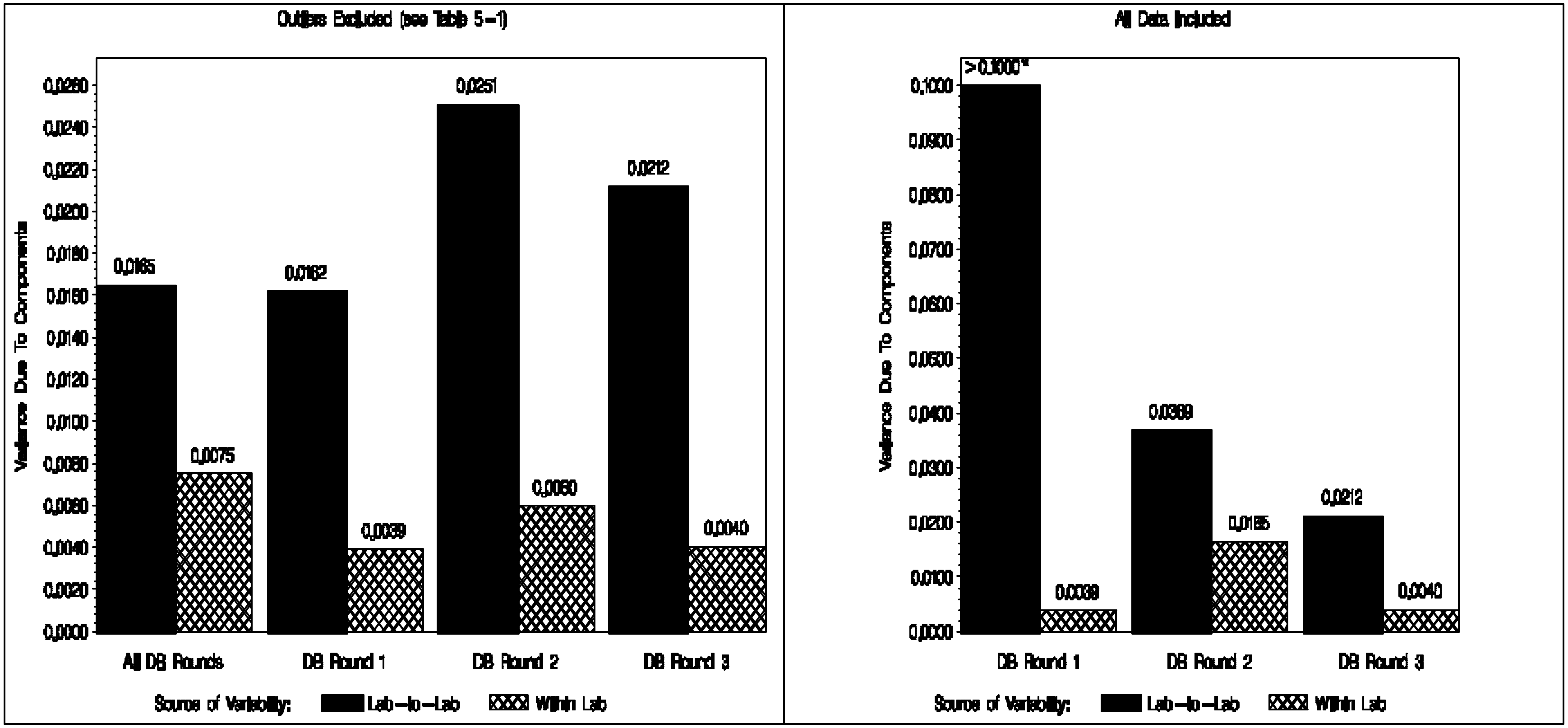

Figure 5-14. Estimates of Lab-to-Lab and Within-Lab Sources of Variability in Reported Lead Amounts Within Mid-Spiked Dust Samples, Based on Analyses Performed With and Without Statistical Outliers Included

Note: Vertical axis is in $(\log(\mu g))^2$.
Note: “All DB Rounds” represents estimates over the entire pilot study. These estimates were calculated only when statistical outliers were removed, as the estimates differed significantly from one DB testing round to another when the outliers were included.

* Value of estimate is 1.7752.

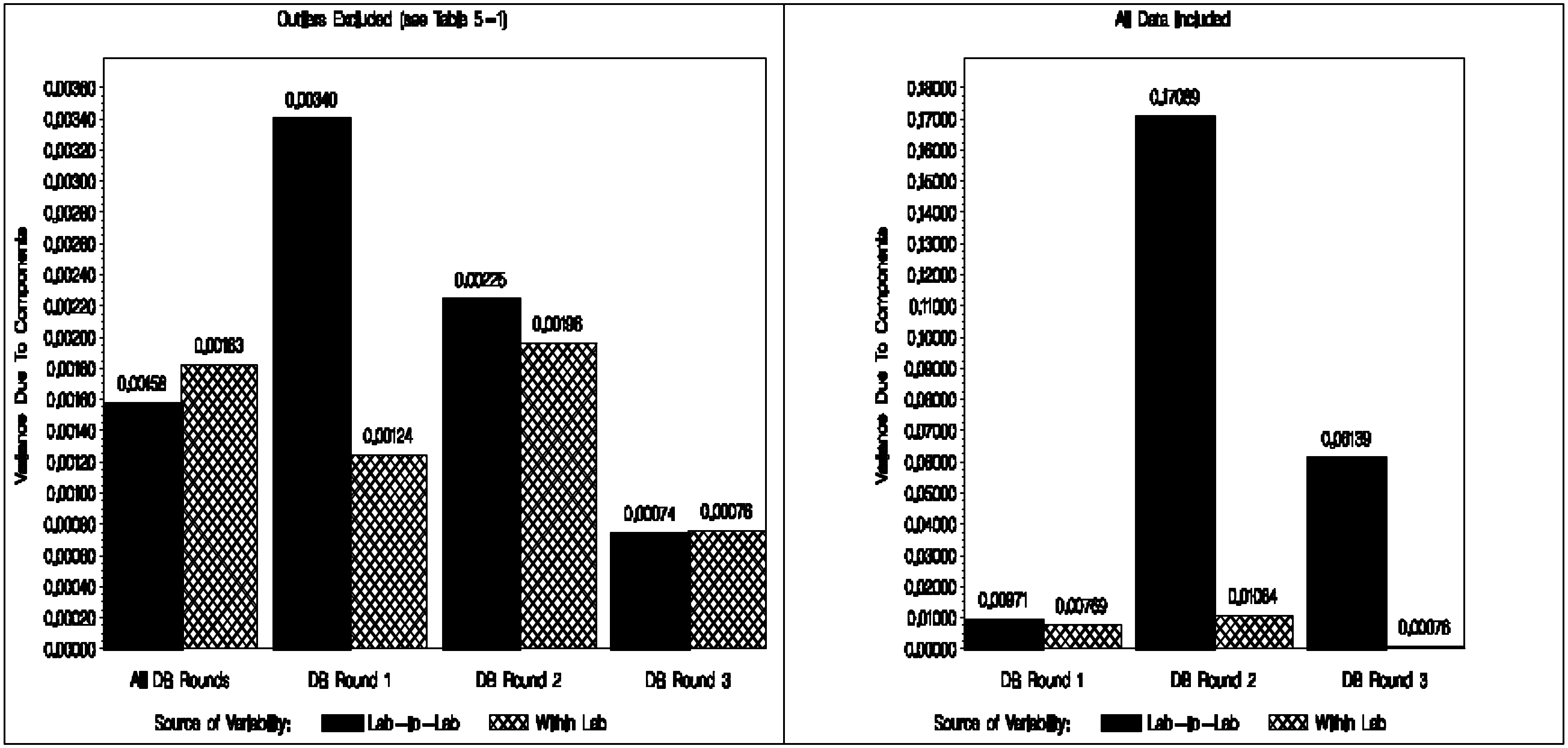

Figure 5-15. Estimates of Lab-to-Lab and Within-Lab Sources of Variability in Reported Lead Amounts Within Spiked Paint Samples, Based on Analyses Performed With and Without Statistical Outliers Included

Note: Vertical axis is in $(\%)^2$.

Note: "All DB Rounds" represents estimates over the entire pilot study. These estimates were calculated only when statistical outliers were removed, as the estimates differed significantly from one DB testing round to another when the outliers were included.

Table 5-12. Estimates of Lab-to-Lab Variability and Within-Lab Variability, Expressed Absolutely and Relative to Total Variability, Associated With Lead Measurements Reported in the Double-Blind Pilot Study, by Sample Type and Testing Round

Sample Type	Testing Round	Lab-to-Lab Variability (% of Total Variability)	Within-Lab Variability (% of Total Variability)	Total Variability
Statistical Outliers Removed (see Table 5-1)				
Low-Spiked Dust Samples	DB Round 1	0.0189 (70.9%)	0.0078 (29.1%)	0.0267
	DB Round 2	0.0274 (84.3%)	0.0051 (15.7%)	0.0325
	DB Round 3	0.0331(85.2%)	0.0057 (14.8%)	0.0388
	Across All Rounds	0.0208 (68.6%)	0.0095 (31.4%)	0.0303
Mid-Spiked Dust Samples	DB Round 1	0.0162 (80.5%)	0.0039 (19.5%)	0.0201
	DB Round 2	0.0251 (80.8%)	0.0060 (19.2%)	0.0311
	DB Round 3	0.0212 (84.0%)	0.0040 (16.0%)	0.0252
	Across All Rounds	0.0165 (68.7%)	0.0075 (31.3%)	0.0240
Spiked Paint Samples	DB Round 1	0.0034 (73.2%)	0.0012 (26.8%)	0.0046
	DB Round 2	0.0022 (53.4%)	0.0020 (46.6%)	0.0042
	DB Round 3	0.00075 (49.5%)	0.00076 (50.5%)	0.00151
	Across All Rounds	0.0016 (46.4%)	0.0018 (53.6%)	0.0034
All Data Included				
Low-Spiked Dust Samples	DB Round 1	1.7075 (99.6%)	0.0074 (0.4%)	1.7149
	DB Round 2	0.0308 (64.7%)	0.0168 (35.3%)	0.0476
	DB Round 3	0.0331 (85.2%)	0.0057 (14.8%)	0.0388
Mid-Spiked Dust Samples	DB Round 1	1.7752 (99.8%)	0.0039 (0.2%)	1.7791
	DB Round 2	0.0369 (69.1%)	0.0165 (30.9%)	0.0534
	DB Round 3	0.0212 (84.0%)	0.0040 (16.0%)	0.0252
Spiked Paint Samples	DB Round 1	0.0097 (55.8%)	0.0077 (44.2%)	0.0174
	DB Round 2	0.1709 (94.0%)	0.0108 (6.0%)	0.1817
	DB Round 3	0.0614 (98.8%)	0.00076 (1.2%)	0.0621

Note: "Across All Rounds" represents estimates over the entire pilot study. These estimates were calculated only when statistical outliers were removed, as for each proficiency-test sample type, the estimates did not differ significantly from one DB testing round to another (at the 0.05 level) when the outliers were removed, but they did differ significantly when the outliers were included.

spiked dust, mid-spiked dust, and spiked paint samples, respectively. Each bar in these charts represents 100% of the total variability. Figures 5-13 through 5-15 present the estimated variance components for low-spiked dust, mid-spiked dust, and spiked paint samples, respectively.

Note that the portion of Figures 5-10 through 5-15 associated with analyses performed while excluding statistical outliers from Table 5-1 include bars for estimates across DB testing rounds (i.e., over the entire pilot study). When statistical outliers were omitted from the analysis, the observed differences in the variance component estimates did not differ significantly (at the 0.05 level) from round-to-round for either component or for any of the three proficiency-test sample types. In these cases, common estimates of lab-to-lab variability and within-lab variability were generated across the entire study, without regard to DB testing rounds, and are included in these figures. However, when statistical outliers were not omitted from the analysis, the variance component estimates differed significantly from one testing round to another ($p < 0.001$), for each type of proficiency-test sample and for both variance components. Thus, variance component estimates over the entire pilot study were not generated when all study data were included in the analysis.

The following additional conclusions could be made from the analyses documented in Figures 5-10 through 5-15 and Table 5-12:

- According to all forms of the model and for each proficiency-test sample type, the overall model-predicted value for the lead measurement in a given DB testing round (represented by the term $\mu+R_j$ in Model (1)) did not differ significantly across testing rounds at the 0.05 level. Furthermore, a predicted value did not differ significantly from its corresponding target level at the 0.05 level. This finding was observed regardless of whether the statistical outliers in Table 5-1 were included or excluded from the analysis. This result implies that general bias in the laboratory-reported measurements on the double-blind proficiency-test samples did not differ significantly across the DB testing rounds (at the 0.05 level), and deviation from their respective target levels (Table 2-1) was not statistically significant overall.

- When all data, including statistical outliers, were included in the analysis, lab-to-lab variability for both low-spiked and mid-spiked dust-wipe samples was greatest in DB Round 1, primarily due to the results for one laboratory (laboratory 07). For paint chip samples, lab-to-lab variability was greatest in DB Rounds 2 and 3, again primarily due to the results for one laboratory (laboratory 09). In each instance, the proportion of total variability associated with lab-to-lab variability exceeded 90%. This percentage decreased considerably when these and other statistical outliers were omitted from the analysis, especially for paint samples (Figure 5-12). Because some laboratories that reported unusually high or low measurements within a DB testing round did so for all samples in that round and with relatively good precision (as noted in Figures 5-1 through 5-3), the presence of statistical outliers highly influenced lab-to-lab variability within a DB testing round.

- When excluding outliers from the analysis, lab-to-lab variability tended to represent approximately 70% of total variability associated with the two types of dust-wipe samples and slightly under 50% of total variability associated with the paint chip samples. These percentages were slightly higher when estimated for only a single testing round, as combining a laboratory's round-to-round variability with its variability associated with analyzing multiple samples within the same testing round contributes to an increased within-laboratory variability estimate.

5.4 COMPARING LAB-TO-LAB VARIABILITY BETWEEN THE ELPAT PROGRAM AND THE DOUBLE-BLIND STUDY

For each of the three sample types included in the double-blind pilot study, results for Round 22 of the ELPAT Program were plotted in Figures 5-1 through 5-3 and listed in Tables 5-6a through 5-8a for the participating laboratories. As each laboratory analyzed only one sample of a given sample type in each testing round within the ELPAT Program, all sources of variability in data from the ELPAT Program are confounded with lab-to-lab variability. Nevertheless, to investigate how statistical acceptance criteria developed for the ELPAT Program may be applied within a double-blind program, it was of interest to compare the variability across laboratories in Round 22 of the ELPAT Program with each testing round of the double-blind pilot study. The approach used to make this statistical comparison was presented in Section 4.4.3.

Initially, for a given double-blind proficiency-test sample type, all sample results reported by a given laboratory within a given DB testing round were averaged. Table 5-13 presents the means and standard deviations of these laboratory averages for each DB testing round (both including and excluding the statistical outliers in Table 5-1), and of the single-sample results reported in Round 22 of the ELPAT Program (and documented in Tables 5-4a through 5-6a). Thus, for the participating laboratories, this table gives an indication of how lab-to-lab variability differs between Round 22 of the ELPAT Program and each round of the double-blind pilot study.

As discussed in Section 4.4.3, Levene's test of homogeneity of variance was used to determine whether the variability estimates presented in Table 5-13 (i.e., the standard deviations) differed significantly across the four testing rounds (the three DB testing rounds and ELPAT Round 22). Significant differences in variability were observed across testing rounds at the 0.05 level only for the two dust-wipe sample types, when the statistical outliers were included in the analysis. This is reflective of the highly-inflated variability observed in DB Round 1 versus the other testing rounds, which resulted from the presence of the statistical outliers. No other incidences of significant differences across testing rounds were observed, despite slightly lower variability estimates occurring in ELPAT Round 22 versus the DB testing rounds.

Table 5-13. Summaries, Calculated Across Laboratories, of Laboratory Average Lead Measurements for the Three Types of Proficiency-Test Samples, Within Each Round of the Double-Blind Pilot Study and in Round 22 of the ELPAT Program

	# Samples Analyzed Per Lab	Summary of Average Sample Result Per Laboratory ------------------------------------ Average (Standard Deviation) (# Laboratories)		
		Low-Spiked Dust (μg Lead)	Mid-Spiked Dust (μg Lead)	Paint (% Lead by Wgt.)
ELPAT Round 22	1	135.5 (10.1) (10)	284.2 (19.8) (10)	0.655 (0.056) (11)
All Double-Blind Pilot Data				
DB Round 1	2 to 6	118.4 (47.2) (9)	250.5 (97.8) (9)	0.622 (0.116) (10)
DB Round 2	2 to 6	134.8 (27.1) (10)	270.9 (50.1) (10)	0.791 (0.421) (10)
DB Round 3	1 to 6	131.1 (23.1) (9)	286.7 (38.7) (9)	0.733 (0.249) (10)
Double-Blind Pilot Data with Outliers Removed (see Table 5-1)				
DB Round 1	1 to 6	132.9 (19.8) (8)	281.2 (35.4) (8)	0.648 (0.061) (10)
DB Round 2	1 to 6	128.8 (21.5) (10)	276.9 (42.3) (10)	0.656 (0.052) (9)
DB Round 3	1 to 6	131.1 (23.1) (9)	286.7 (38.7) (9)	0.655 (0.032) (9)

Note: Statistics in this table for ELPAT Round 22, are based on one sample analyzed per laboratory and considers only those laboratories involved in the double-blind pilot study. Statistics in this table for the double-blind (DB) pilot study are based on averages of multiple samples analyzed per laboratory.

Caution must be exercised when interpreting the results in Table 5-13 and the results of Levene's test. For each laboratory, the averages in this table for the DB testing rounds are calculated from laboratory averages of from up to six results each, while averages for Round 22 of the ELPAT Program are calculated from individual sample results. The average of multiple observations from a common distribution has lower variability than any one observation from this distribution (Snedecor and Cochran, 1989), and so, the two sets of data would be expected to have different underlying variability. To extend this point further, the double-blind averages should have a lower standard error than the ELPAT Program average, if the double-blind data originate from the same distribution as the ELPAT Program data and a constant number of laboratories is assumed. However, in most instances, the opposite is seen in Table 5-13. Regardless of whether statistical outliers were included or not, the standard deviations in Table 5-13 for the two dust sample types were higher (by at least 79%) in each of the double-blind pilot testing rounds compared to Round 22 of the ELPAT Program. Only for paint sample results in Rounds 2 and 3 when outliers were excluded were the standard deviations of the laboratory averages below what was observed in Round 22 of the ELPAT Program.

Therefore, while Table 5-13 implies that averages in the DB testing rounds tended to vary more considerably across laboratories than did the individual sample results within Round 22 of the ELPAT Program, the extent that this lab-to-lab variation differed across testing rounds was not necessarily statistically significant. However, as the table summarized averages of multiple sample results for each laboratory in the DB testing rounds, while single-sample results were summarized from ELPAT Round 22, one expected to see lower variability in the DB testing rounds if, in fact, the data for the double-blind and ELPAT Program testing rounds originated from the same underlying distribution. Therefore, even if a double-blind program evaluates a laboratory based on an average result across multiple samples analyzed by the laboratory (of the same type), rather than on the result of analyzing an individual sample, the evaluation criteria should consider that the results of double-blind testing may have greater lab-to-lab variability compared to the results of single-blind testing.

This page left blank intentionally.

6.0 QUALITY ASSURANCE

A successful proficiency-test program must apply quality assurance procedures to ensure the overall integrity of the proficiency-test samples throughout the course of the program, from preparation to analysis. In addition, the program must ensure that analytical results involving the proficiency-test samples are reported accurately. This chapter discusses how these issues were addressed in this double-blind pilot study. More details on quality assurance issues in this pilot study are included in Appendix A.

6.1 SAMPLE FABRICATION AND TRANSFER

As discussed below, the dust and paint source materials used in the double-blind pilot study were obtained as part of the ELPAT Program. The proficiency-test samples were prepared from these materials within the ELPAT Program. Some of the procedures used to prepare the proficiency-test samples are discussed below, with an emphasis on quality assurance practices.

6.1.1 Obtaining and Preparing Bulk Source Material

The paint and household dust source material used to prepare the proficiency-test samples were collected in the ELPAT Program following the procedures set forth in *Standard Operating Procedure for Source Material Collection* (Appendix C of RTI, 1994). A network of contractors associated with abatement and risk assessment projects in public housing, military housing, and private dwelling units contributes bulk paint source material for use in the ELPAT Program, while dust sample material is obtained from vacuum bags in households conducting normal cleaning routines, from HEPA vacuums used in post-abatement cleaning efforts, and from street sweeping. This material is then classified according to lead content. Specifically, the low-lead dust used in this study came from a Milwaukee (WI) exposure intervention program, the medium-lead dust came from a North Carolina household, and the paint came from an old hospital in Raleigh, NC. See Section 2 of Appendix A for additional information on material selection.

In the ELPAT Program, a given batch of source material is homogenized with respect to lead concentration and particle size distribution. The method for preparing the paint source material is detailed in *Standard Operating Procedure for Preparation of Lead-In-Paint Proficiency Analytical Testing Material* (Appendix C of RTI, 1994) and summarized in Section 4.3.1 of RTI, 1994. The method for preparing the dust source material is detailed in *Standard Operating Procedure for Preparation of Proficiency Analytical Testing Material for Lead in Dust* (Appendix D of RTI, 1994) and is summarized in Section 4.3.2 of RTI, 1994.

6.1.2 Preparing Proficiency-Test Samples

The dust and paint proficiency-test samples were prepared within the ELPAT Program. For the double-blind pilot study, 180 proficiency-test samples of each sample type were prepared (Table 2-1 of Section 2.2). Additional samples at the rate of 5% per sample type were also prepared to act as QC samples in the final verification process, which verified the lead content in the proficiency-test samples, as described in Section 3.3 of Appendix A. The methods used to prepare the proficiency-test samples are detailed in Sections 4.3.1 and 4.3.2 of RTI, 1994, and in Section 3 of Appendix A.

Each dust proficiency-test sample consisted of a PaceWipe™ with 0.1 (±0.0005) grams of the appropriate dust source material, as described in Appendix D of RTI, 1994[3]. The PaceWipe™ is preferred in the ELPAT Program as it contains no detectable background lead when using flame atomic absorption spectrometry, such as NIOSH Method 7082. As presented in Section 4.1 of RTI (Appendix A), analyses of 13 blank PaceWipes™ each reported <0.001 mg lead/wipe. In addition, the PaceWipe™ has a consistent moisture level and is covered by a solution that tends to retard molding over time when stored under proper environmental conditions.

Paint proficiency-test samples consist of one-gram aliquots of the paint source material.

When dust (at each lead level) and paint proficiency-test samples were tested in duplicate, the difference in results agreed within 5% of the initial result. Potential matrix interferences were evaluated for both dust-wipe and paint-chip proficiency-test materials by evaluating the recovery of lead spiked into replicate samples before analysis. The lead recoveries of the spiked samples were within the target recovery range of 90-110%. These results are discussed in Section 4.1 of Appendix A.

Once prepared, the proficiency-test samples were placed into individual plastic scintillation vials, capped, and stored until they were ready to be transferred to client-supplied sample containers (Section 2.4.1). Dust samples were stored in a 4°C cold room. All vials were stored according to the type of sample (dust, paint) and the level of lead in the sample.

6.1.3 Verifying That Client-Supplied Sample Containers Are Uncontaminated

From each batch of sample containers received from a client, one container was used in a lead-background test to verify that the batch of containers were uncontaminated. This test involved swabbing the interior of the container with a PaceWipe™, then analyzing the wipe for lead contamination (Section 4.3 of Appendix A). While a default sample container (e.g., plastic centrifuge tubes different from the containers used in the ELPAT Program) would have been used in place of a client's sample containers if they were found to be contaminated, this corrective action was not necessary (i.e., all analysis results were <0.001 mg lead/container).

[3] While RTI, 1994, indicates that Whatman™ No. 40 filters are used, the PaceWipe™ was used in this pilot study and in the ELPAT Program.

6.1.4 Transferring Samples to Client-Supplied Sample Containers

To verify the quantitative transfer of dust-wipe samples from the plastic scintillation vials to the client-supplied sample containers, seven blank PaceWipes™ were spiked with NIST Standard Reference Material 2711. The samples were digested and analyzed for lead, as detailed in Appendix D of RTI 1994, and in Section 4.2 of Appendix A. The recovery percentage averaged 83.3% (±1.6%) for these seven samples, compared to an 85% nominal recovery percentage for the SRM and the recovery percentages of 82.6% - 86.0% recorded in Round 22 of the ELPAT Program.

To investigate whether PaceWipes™ could become contaminated with lead as a result of the sample transfer process, three blank PaceWipes™ were stored in plastic scintillation vials for two hours and stored in a 4°C cold room, transferred to the default centrifuge tube, and removed for analysis. The blank recoveries for these three samples were each <0.001 mg lead/wipe (Section 4.1 of Appendix A).

6.2 DATA MANAGEMENT AND SAMPLE TRACKING

Data collected in this double-blind pilot study were generated by multiple sources specified within Figure 2-1 of Section 2.0. This section describes the data management procedures that were used in this pilot study and the methods used for sample tracking through the study.

6.2.1 Types of Data

The majority of data in the pilot study were the quantitative and tracking data that were taken from the Sample Tracking and Analysis Report Forms (Appendix C). Other types of data included the analytical results for the participating laboratories from Round 22 of the ELPAT Program (provided by the proficiency-testing service), qualitative information on the participating clients (used to select clients in Chapter 3), and any feedback that the clients (or laboratories) had as a result of their participation in the study.

Information on the recruited clients was recorded during the recruitment process onto copies of the telephone recruitment script in Appendix B. While most of this information was used to determine the eligibility of the client for this pilot study (Sections 3.2 and 3.3) and to determine which laboratories were testing proficiency-test samples, selected information was used for correspondence and sample shipment throughout the study.

The proficiency-testing service recorded the PTS Sample ID (Section 5.1 of Appendix A), sample type, sample weight, sample shipped date, and sample received date on the Sample Tracking and Analysis Report Forms. They also prepared the report in Appendix A detailing the results of verification testing and other information on the proficiency-test samples such as sample-to-sample variation determined within the sample preparation stage.

The participating clients recorded their own Client Sample IDs for cross-reference, dates of sample shipment and receipt of analysis results, and the analysis results for the proficiency-test samples on the Sample Tracking and Analysis Report Forms. Clients also could report protocol violations and quality assurance issues to the proficiency-testing service when necessary.

6.2.2 Data Storage and Transfer

Throughout the pilot study, the following hardcopy documents were stored by the organizations responsible for their completion and reporting: completed telephone recruitment scripts, copies of the Sample Tracking and Analysis Report Forms that were sent to the clients with the sample shipments, and copies of these same forms as received from the clients via fax as they shipped the samples to the laboratories and as they received the analytical results.

Section 2.5 discusses how data were reported from the various organizations involved in the pilot study and how these data were stored electronically.

6.2.3 Sample Identification

Identifications were placed on proficiency-test samples at two distinct points in the double-blind pilot study: when samples were prepared by the proficiency-testing service ("PTS Sample IDs") and when samples were incorporated into regular field sample batches by the participating clients ("Client Sample IDs").

The method that the proficiency-testing service used to specify PTS Sample IDs is discussed in Section 5.1 of Appendix A. These sample IDs were placed onto labels which were affixed to plastic bags. Then, the appropriate sample containers were placed in their appropriate plastic bags. The proficiency-testing service recorded the PTS Sample IDs on Sample Tracking and Analysis Report Forms and included the forms with the samples when shipping to the clients.

When the proficiency-test samples were received by the clients, they assigned Client Sample IDs to the samples when placing them in a batch for shipment to the laboratory. The identifiers were assigned in a manner that the laboratory could not distinguish the proficiency-test samples from the other field samples in the batch based on its ID or label.

After recording the Client Sample IDs in Section B of the Sample Tracking and Analysis Report Form next to the PTS Sample ID, the client removed the proficiency-test sample container from the plastic bag and affixed a label containing the Client Sample ID onto the container. The type of label used by the client, the manner of recording Client Sample IDs to the labels, and the manner of affixing the labels to the sample containers were consistent across all samples in the batch to ensure blindness of this pilot study.

6.3 DATA QUALITY CHECKING

The proficiency-testing service performed some verification of hand-entered data before releasing the final spreadsheet of analytical results for statistical analysis. The proficiency-testing service also provided the originals of the Sample Tracking and Analysis Report Forms and laboratory reporting forms that the participating clients provided. Once the organization performing the statistical analysis received these materials, they performed a 100% verification of the data in the spreadsheet for each testing round, comparing the recorded results with what was recorded on the forms. Any deviation from the forms was reported back to the proficiency-testing service for verification. This process was completed before preparing final versions of the results presented in Chapter 5.

The organization performing the statistical analysis also notified the proficiency-testing service of any results that appeared to be extreme (i.e., unusually high or low) relative to other results for the given sample type, or relative to the target lead level as determined from the reference labs in Round 22 of the ELPAT Program. The proficiency-testing service investigated the correctness of these extreme data values by contacting the laboratories that analyzed the samples in question to obtain the analytical results as reported on the laboratory report forms, and/or contacting the clients associated with these samples to verify that they reported the results correctly and in the proper units. The proficiency-testing service provided all information obtained in this investigation to the organization performing the statistical analysis. Any necessary data corrections that were identified in this investigation into extreme data values were made prior to generating the final data summaries and analyses presented in Chapter 5.

This page left blank intentionally.

7.0 CONCLUSIONS AND RECOMMENDATIONS

The following findings and conclusions were made from this double-blind proficiency-testing pilot study and from the summaries and analyses of the data generated by this study:

Design and participant issues

- It can be difficult to get clients of laboratories (e.g., risk assessors) to cooperate in a double-blind proficiency-testing program (through receiving and distributing proficiency-test samples and reporting back the results of the analysis of these samples) without providing some kind of incentive or reimbursement.

- Reimbursement for client participation in a double-blind program would require funding and a small amount of management staff effort to administer the funds.

- It is apparent that some laboratory clients may not recognize the benefits that a double-blind proficiency-test program would have for them, over and above current proficiency-testing programs.

- Many laboratory clients are small operations that do not have the cash flow or staff to contribute resources to a double-blind program, making their participation a hardship to them. In some cases, materials (e.g., sample containers, PaceWipes™) were provided to clients in this pilot study to ensure their participation.

- Simple, yet explicit, instructions are necessary for the clients to ensure proper storage, handling, and identification of proficiency-test samples while in their control. These factors can affect the ongoing integrity of the samples.

- Frequently, clients go out of business, reorganize, change their organizational identity and/or mission, and change their telephone numbers and staff. In addition, points of contact are frequently unavailable when needed (typically in the field). This hinders the ability of clients to give a long-term commitment to a double-blind program, as well as the ability of a proficiency-testing service to contact participating clients during the course of a double-blind program.

- Several clients participating in this pilot study (and many who were attempted to be recruited) had a low monthly volume of field samples, making it difficult to generate a batch containing field samples and proficiency-test samples within a month of receiving the proficiency-test samples. This problem may be especially acute in time periods when less environmental sampling occurs (e.g., winter months).

- The finely-ground nature of paint-chip proficiency-test samples, along with the absence of substrate particles, make these samples easily distinguishable from field samples. Further research may be necessary to reformulate the physical characteristics of paint-chip proficiency-test samples.

- As soil is recognized as a potential source of lead in indoor dust, soil sampling can play an important role in a risk assessment. However, while field soil samples typically exceed 10 grams, the proficiency-test soil samples in the ELPAT Program are typically less than 5 grams. More study is needed to determine how soil proficiency-test samples can be prepared for a double-blind program.

- Currently, the use of PaceWipes™ or a similar hand-towelette in preparing dust-wipe proficiency-test samples makes it easy to distinguish them from field samples that consist of baby wipes, which many risk assessors use. This can hinder the participation of clients who use baby wipes. However, it is expected that laboratory requests to use the smaller, thinner towelettes rather than baby wipes will result in more clients adopting towelettes for dust sampling.

- Future double-blind programs need to take into account situations where proficiency-test sample results are reported as a lead amount per "unit area." When blindly including proficiency-test samples within a batch of regular field samples, the clients fabricated field sampling information associated with the proficiency-test samples to aid in their disguise. This fabricated information included sampled areas associated with the samples. This posed a problem when clients reported back the results of the proficiency-test samples in terms of a lead amount per unit area. In order to convert this to a strict lead amount, it was necessary to verify the "area" that the client fabricated for this sample.

- Page 5-33 of the 1995 HUD Guidelines recommends that risk assessors use double-blinding techniques on their laboratories, where risk assessors obtain spiked dust-wipe samples (in the range of 50-300 µg/wipe) from laboratories and insert them into their field sample batches for analysis (at 1 spiked wipe per 50 samples). However, this is a recommendation and not a requirement. Except when double-blinding is mandatory (e.g., certain government programs such as the HUD Grantees program), the added resources that double-blinding requires on the part of the risk assessor often keep some from performing (and understanding) double-blinding on a voluntary basis. This issue must be addressed when determining the feasibility of a double-blind proficiency testing program.

Laboratory performance

- While all data for Round 22 of the ELPAT Program fell within the round's acceptance limits for the 12 participating laboratories in the double-blind pilot study, 11% of the double-blind pilot data for these laboratories exceeded these limits. This suggests that for some (but not necessarily all) of the participating laboratories and for each sample type, the double-blind pilot study data are more likely than single-blind study data to exceed the acceptance limits determined from data within the (single-blind) ELPAT Program.

- For each proficiency-test sample type, the overall average measurement reported by the laboratories within a double-blind testing round did not differ significantly across the three testing rounds (at the 0.05 level), and deviation of this average from the target level associated with the proficiency-test sample type was not statistically significant overall.

- Some statistical evidence exists that the variability in log-transformed dust-wipe proficiency-test sample measurements differs significantly (at the 0.05 level) across double-blind testing rounds, primarily due to the presence of unusually large or small data values. Variability in the dust-wipe measures also tended to be higher in the double-blind testing rounds compared to Round 22 of the ELPAT Program. These observations were less evident for the paint chip proficiency-test sample measures.

- Analysis of the double-blind pilot study data characterized variability into two components: lab-to-lab variability and within-lab variability. For both of these components, when statistical outliers were omitted from the analysis, the observed differences in the variance estimates did not differ significantly (at the 0.05 level) from round-to-round for any of the three proficiency-test sample types. This finding, along with the conclusion made in the previous bullet, suggests that general laboratory performance did not differ significantly across the three double-blind testing rounds.

- When excluding outliers from the analysis, lab-to-lab variability tended to represent approximately 70% of total variability associated with the two types of dust-wipe samples and slightly under 50% of total variability associated with the paint chip samples. Thus, lab-to-lab variability constituted a greater percentage of total variability for the dust-wipe proficiency-test samples than for the paint chip samples. These percentages were slightly higher when calculated within each double-blind testing round.

- When statistical outliers were not excluded from the analysis, the estimates of both lab-to-lab variability and within-lab variability differed significantly from one testing round to another ($p < 0.001$), for each type of proficiency-test sample. However, as those laboratories that occasionally reported unusually high or low results did so for multiple samples of a given type within a testing round, the presence of statistical outliers affected lab-to-lab variability considerably more than within-lab variability, often resulting in lab-to-lab variability representing over 90% of total variability in the given testing round.

- When calculating laboratory averages within each double-blind testing round (disregarding statistical outliers) and the single-sample results from Round 22 of the ELPAT Program for the same group of laboratories and the same proficiency-test sample types, no statistically significant differences (at the 0.05 level) were observed in the variability of these data across the double-blind testing rounds and Round 22 of the ELPAT Program. However, the observed variability associated with the double-blind laboratory averages was slightly higher than the observed variability associated with the single-sample results for these laboratories in Round 22 of the ELPAT Program. Because statistical theory specifies that averages have lower variability than the data entering into their calculation, this finding suggests that for these laboratories and proficiency-test sample types, the results of double-blind testing may have higher variability compared to the results of single-blind testing.

- The data for the group of laboratories participating in this pilot study suggest that additional variability may be present in double-blind testing data compared to single-blind testing data. However, while such a finding may influence how acceptance ranges in a double-blind program are determined, the criteria for determining acceptance in a double-blind program should not be relaxed simply because laboratories may be more likely to exhibit reduced performance compared to within a single-blind program. Instead, the criteria should address a laboratory's typical performance level in the test setting.

The following <u>recommendations</u> can be made as a result of conducting this study:

- An education strategy is needed for laboratory clients to recognize the benefits of a double-blind program.

- Consideration should be given on whether clients should be reimbursed for costs associated with proficiency-test sample analyses in a double-blind program. If the decision to reimburse is approved, sources of funding must be identified, such as charging laboratories a fee for participating in a double-blind program.

- In a double-blind proficiency-testing program, it may be beneficial to have proficiency-test samples placed in small batches rather than large batches. If the proficiency-test sample results are considerably different from those for other samples within the batch (e.g., are very high), then some laboratories may suspect that the samples are some kind of reference material. This is more likely to happen in large batches than in small batches, especially if the lead content is consistent from sample to sample within a batch.

- Statistically-based laboratory performance criteria in a double-blind program may need to consider that lab-to-lab variability in proficiency-test sample results may naturally differ between a double-blind setting and a single-blind setting, as some laboratories may perform differently when aware of analyzing proficiency-test samples.

- If statistical evaluation criteria in a double-blind program will be made based on individual sample results (as is done in the single-blind program), then within-laboratory variability in these results should be considered, in addition to lab-to-lab variability.

- Risk assessors can use several different types and brands of wipes for collecting dust samples for lead analysis. Therefore, it is necessary to work with laboratories, their clients, and other interested agencies to standardize the type of dust-wipe that should used in lead inspections and risk assessments, so that the same type of wipe can be used to develop proficiency-test samples.

- Further research should be considered to develop more appropriate paint materials that can be used in proficiency-test samples and that more closely resemble paint samples collected in the field than the finely-ground material used in this pilot study. A similar recommendation can be made for soil samples, which were not considered in this pilot study.

- When laboratory proficiency-test results are suspected to have been reported inaccurately, and it is determined that the laboratory client did not contribute to the inaccuracy, it may be necessary for the proficiency testing service to contact the laboratories to resolve any issues contributing to the inaccuracy, such as by reviewing laboratory QA/QC results for the given batch (e.g., results of analyzing laboratory control samples), and to determine whether the client's field sample results within the same batch may have been compromised in any way.

This page left blank intentionally.

8.0 REFERENCES

"EPA Seeks Lab Clients for 'Double Blind' Test Program." *Lead Detection and Abatement Contractor*, January 1998,'Newsline' column, page 2.

NIOSH (1994) *Laboratory Evaluations and Performance Reports for the Proficiency Analytical Testing (PAT) and Environmental Lead Proficiency Analytical Testing (ELPAT) Programs.* Prepared by the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention, November, 1994.

Rosner, B. (1983) "Percentage Points for a Generalized ESD Many-Outlier Procedure." *Technometrics* 25(2):165-172.

RTI (1994) *Environmental Lead Proficiency Analytical Testing Program: Quality Assurance Project Plan, Revision No. 1*. Prepared by the Center for Environmental Measurements and Quality Assurance, Research Triangle Institute, for the American Industrial Hygiene Association, 3 May 1994.

Schlecht, P.C., Groff, J.H., Feng, A., and Song, R. (1996) "Laboratory and Analysis Method Performance of Lead Measurements in Paint Chips, Soils, and Dusts." *American Industrial Hygiene Association Journal* 57:1035-1043.

Snedecor, G.W., and Cochran, W.G. (1989) *Statistical Methods, Eighth Edition*. Ames, IA: Iowa State University Press.

USEPA (1995) "A Field Test of Lead-Based Paint Testing Technologies: Summary Report." Office of Prevention, Pesticides, and Toxic Substances, U.S. Environmental Protection Agency, EPA 747-R-95-002a, May 1995.

USHUD (1995) "Guidelines for the Evaluation and Control of Lead-Based Paint Hazards in Housing." Office of Lead Hazard Control, U.S. Department of Housing and Urban Development. HUD-1539-LBP.

APPENDIX A

Report on the Preparation and Distribution of Samples for the Double-Blind Proficiency Testing Pilot in the National Lead Laboratory Accreditation Program

Prepared for

Fred I. Grunder
American Industrial Hygiene Association
2700 Prosperity Ave., Suite 250
Fairfax, VA 22031

Prepared by

Laura L. Hodson, David A. Binstock and Ann R.Turner
Environmental Chemistry Department
Center for Environmental Measurements and Quality Assurance

June 1998

SECTION 1.0
INTRODUCTION

1.1 PROJECT OVERVIEW

This report describes the preparation and distribution of lead-containing dust and paint samples for a pilot version of a double-blind proficiency testing program. The double-blind program is currently being considered for inclusion within the National Lead Laboratory Accreditation Program (NLLAP) operated by the U.S. Environmental Protection Agency to test the ability of analytical laboratories to analyze dust, soil, and paint chip samples for lead content.

The current laboratory proficiency testing program within NLLAP is the Environmental Lead Proficiency Analytical Testing (ELPAT) program which is administered by the American Industrial Hygiene Association (AIHA). Since 1992, AIHA has contracted with Research Triangle Institute (RTI) to produce lead-containing paint, soil and dust wipe samples for distribution to laboratories wishing to participate in the ELPAT program. The ELPAT program is a single-blind proficiency testing program, so while the participating laboratories are unaware of the amount of lead in samples that are part of the program's performance evaluation, they are aware of when they are analyzing such samples.

This double-blind study will have the field inspectors (clients) receiving lead-in-dust and lead-in-paint samples from RTI, and incorporating these proficiency-testing samples as blind samples within batches of field samples submitted by the clients of the laboratories. Since these proficiency samples would not be identified as such in these batches, laboratories would ideally treat these samples as routine field samples.

To accomplish such a task, RTI prepared additional Round 22 ELPAT materials, and mailed these samples to participating clients. Samples were provided to the clients three times over a three month period. It was the clients responsibility to report the laboratory analysis results back to AIHA.

1.2 PROJECT ORGANIZATION AND MANAGEMENT STRUCTURE

The double-blind pilot program was sponsored by the Office of Pollution Prevention and Toxics of the U.S. EPA (EPA/OPPT) under Work Assignment 3-30 of EPA Contract Number 68-D5-0008. Mr. John Scalara was the EPA Work Assignment Manager. Three contractors were involved in the design and conduct of this program. Battelle is under contract by the EPA, AIHA was a subcontractor to Battelle and RTI was a subcontractor to AIHA.

Battelle had responsibility for the establishment of the program design, preparation of the Quality Assurance Project Plan, recruiting participants, establishing statistical performance criteria, preparing SOPs, letters and forms to be used in the pilot study, and preparing a final report for EPA.

AIHA was responsible for serving as the proficiency testing service, acting as the client's primary contact, collecting the analytical results and providing them to Battelle, and providing data on the ELPAT Round materials.

RTI was responsible for preparing and storing the proficiency test materials used in the double-blind study, receiving the clients sample containers and testing these containers for any possible lead contamination, conducting verification procedures on sample transfers among storage devices and centrifuge tubes, distributing samples to clients using the client-supplied containers, (or a default centrifuge tube if the client did not supply a container), coordinating the method for ensuring blindness when the clients incorporate these samples with their regular field samples, and reporting information on the test samples that is necessary for determining laboratory performance.

This report describes the preparation and distribution of the double-blind samples.

SECTION 2.0
PROFICIENCY TESTING MATERIAL SELECTION

2.1 MATERIAL SOURCES

The low-lead dust used in this study came from a Milwaukee exposure intervention project and the medium-lead dust came from a North Carolina household. The paint came from an old hospital in Raleigh, NC. The double-blind samples were prepared from the same bulk processed materials as those used for Round 22 in the ELPAT program. The double-blind samples were prepared immediately preceding Round 22.

2.2 SCREENING ANALYSIS

The dusts were sent to Neutron Products, Inc., in Dickerson, MD for sterilization by gamma irradiation and then returned to RTI. Upon receipt at RTI, the bags of sterilized raw dust were sieved using a 250-um sieve, and three 0.1 g aliquots were taken from the sieved dust and subjected to analysis using microwave/acid digestion and inductively coupled plasma emission spectroscopy[1]. This concentration was the screening value for each dust.

The paint was ground using a ball-mill jar, and three 0.1 g aliquots were subjected to the lead analysis as described for the dust samples. The lead concentration was the screening value for the paint.

2.3 TARGET CONCENTRATION

The QAPP for the double-blind project (written by Dr. Bob Lordo et al., of Battelle, Feb. 1998) specified that the dust-wipe proficiency-test samples will have lead levels in each of the following ranges : 70-120 *u*g/wipe (Low-level dust); and 200-600 *u*g/wipe (Mid-level dust); and that the paint chip sample be in the range of 0.2-1.2 percent lead. Samples were then selected from the Round 22 ELPAT proficiency samples to achieve these target values. The following ELPAT samples were selected: 22W2 (Low-level dust), 22W3 (Mid-level dust), and 22P4 (Paint).

SECTION 3.0
PREPARATION OF DOUBLE-BLIND MATERIALS

3.1 PRELIMINARY VERIFICATION

The bulk dust samples, having been sieved to 250 um for screening analysis, were then passed through a 150 *F*m sieve using the Ro-Tap apparatus in preparation for preliminary verification analysis. The paint having been previously ground using the ball mill jar was passed through a 125 Fm sieve.

Five 0.1 g grab samples of each dust and paint were then taken manually from each batch of the processed materials, and analyzed using the microwave/ nitric acid, hydrochloric acid extraction and measurement by plasma emission spectroscopy. The preliminary verification values as presented in Table 1 came within the acceptable range of the target concentrations.

Table 1
Screening and Preliminary Verification

Sample	Target Value	ELPAT Number	Screening Value	Preliminary Verification	Source
Low-level Dust	70 -120 ug/wipe	22W2	1180 Fg/g	118 ± 5.04 Fg/wipe	Milwaukee, WI Intervention Program
Mid-level Dust	200 - 600 ug/wipe	22W3	2270 Fg/g	243 ± 8.44 Fg/wipe	North Carolina Household
Paint	0.2 - 1.2 %	22P4	0.642 %	0.602 ± 0.044 %	Raleigh, NC Hospital

3.2 SAMPLE LOADING

PaceWipes™ were loaded with 0.1000 ± 0.0005 g portions of dust. Only one analytical balance was used to weigh the dust onto the PaceWipes™ and a single 0.1 g Class 1 weight was used for a daily calibration check. The dust jar was first tumbled, then allowed to settle briefly and the container was opened. Material was taken using a spatula and transferred to the tared weighing paper. If more material was needed, it was taken from the bulk container and added to the material on the paper. If excess (>0.1005 g) was placed on the weighing paper, it was carefully removed with the tip of the spatula and discarded. A PaceWipe™ was prepared for receiving the dust by opening the foil pouch, removing the wet folded wipe and squeezing the excess moisture out by hand over a trash can.

The wipe was then unfolded and briefly set on a Kimwipe to soak up excess moisture. The PaceWipe™ was then transferred to a flat plastic board to await the dust. The weighing paper containing the pre-weighed dust was then removed from the balance and the dust gently tapped out onto the PaceWipe™ . The wipe was then folded and placed in a capped, plastic scintillation vial. All vials containing the spiked wipes were stored in a cold room (40E F) as a secondary means of retarding mold growth until shipment.

The paint was riffled out into 1.0 gram aliquots using a spinning riffler. A 20 gram portion was introduced into the hopper of the riffler, and the paint was slowly vibrated down a chute leading from the hopper to 20 trays slowly turning under the end of the chute. After all of the 20 gram portion was split into 20 samples, the trays were removed and the paint samples transferred to plastic scintillation vials. The process was repeated 9 times to obtain 180 samples.

3.3 FINAL VERIFICATION

After the vials were filled with the appropriate dust or paint samples, they were returned to their positions in divided boxes holding 10 rows of 10 vials each. Samples were selected for final verification across the entire set of samples at a rate of 5% of the total number of samples. One dust or paint sample was selected at random from each batch of 20 samples, for a total of 9 samples from each set of 180 dusts and paint. The final verification values are presented in Table 2. Values within 20% of the target values were achieved.

Homogeneity of the samples is indicated by the relative standard deviations (RSD). The RSDs based upon analysis of 9 samples met the goal of a RSD less than 10%.

Table 2
Final Verification

Sample	Final Verification (n=9)	Relative Standard Deviation (RSD)
Low-level Dust	101 ± 5.29 Fg/ wipe	5.24
Mid-level Dust	215 ± 10.4 Fg/wipe	4.86
Paint	0.658 ± 0.029 %	4.40

SECTION 4.0
QUALITY ASSURANCE

4.1 QUALITY CONTROL OPERATIONS

Quality control/quality assurance is an essential component of the ELPAT program, and was continued for the double-blind study. An earlier study to determine the blank value of the PaceWipe™ (Lot number 1296-01) showed that 13 blanks all contained <0.001 mg lead/wipe. Three more blank wipes which had been placed in plastic scintillation vials for two hours and stored at 40 °F prior to removal and transfer to the centrifuge tube, were subsequently analyzed. The blank recoveries for these three samples were also <0.001 mg lead/wipe.

A duplicate analysis was conducted for each of the dusts and paint. The duplicates agreed within 5 %, as shown in Table 3. Spike solutions were prepared from a 1000-*u*g/ml stock solution of $Pb(NO_3)_2$ obtained from PE Pure, Atomic Spectroscopy Std and added to the dust samples prior to digestion. One ml of a 50 *u*g/ml Pb solution was added to the low-level dust and one ml of a 100 *u*g/ml Pb solution was added to the mid-level dust prior to digestion. The paint solution concentration was too high to effectively add a spike solution prior to digestion; therefore one ml of a 50 *u*g/ml spike solution was added to the diluted paint solution following digestion. The spike recoveries for the dusts and paint were within the goal of 90-110, as shown in Table 4.

Table 3
Duplicate Analysis

Sample	Vial ID	Initial	Repeat	Percent Difference
Low-level Dust	462/482	104 Fg/g	102 Fg/g	1.9 %
Mid-level Dust	471/491	210 Fg/g	218 Fg/g	3.8 %
Paint	465	0.676 %	0.649 %	4.4 %

Table 4
Results of Spike Analysis

Sample	Added Amount (Fg)	Amount Recovered (Fg)	% Recovery	Unspiked Amount (Fg)
Low-level Dust	50	45.6	91.2	101
Mid-level Dust	100	92.4	92.4	214
Paint	50	50	100.0	32.0

4.2 QUANTIFICATION OF DUST-WIPE SAMPLE TRANSFER

To verify the quantitative transfer of dust-wipe samples from the scintillation vials to the client-supplied sample containers, seven aliquots of NIST SRM 2711 were weighed and transferred to the PaceWipes™ as described for the dust loading. These scintillation vials were placed in the cold room overnight, then the wipes were transferred from the scintillation vials to the centrifuge tubes used for the analysis. The wipes were then digested, analyzed and the results compared to the historic recovery of lead from SRM 2711, which is nominally 85%. The recovery of these seven wipes was 83.3 ± 1.6 % which can be compared to the Round 22 SRM 2711 recovery of 82.6 - 86.0%.

4.3 CLEANLINESS DETERMINATION OF CLIENT-SUPPLIED CONTAINERS

Each client supplied an extra container which was tested for lead contamination by swabbing the interior of the container with a PaceWipe™. These PaceWipes were subsequently analyzed for lead contamination. As can be seen on Table 5, all containers contained <0.001 mg lead/container.

Table 5
Cleanliness of Client-Supplied Containers

Client ID	Container Type	Lead from Container
A	Glass Jars	<0.001 mg
	Centrifuge Tubes	<0.001 mg
B	Plastic Bags	<0.001 mg
C	Plastic Bags	<0.001 mg
D	Plastic Bags	<0.001 mg
E	Centrifuge Tubes	<0.001 mg
F	Plastic Bags	<0.001 mg
G	Centrifuge Tubes	<0.001 mg
H	RTI Centrifuge Tubes	<0.001 mg
I	Centrifuge Tubes	<0.001 mg
J	Plastic Bags	<0.001 mg
K	Plastic Bags	<0.001 mg
L	Centrifuge Tubes	<0.001 mg

SECTION 5.0
PACKAGING AND DISTRIBUTION

5.1 SAMPLE IDENTIFICATION

RTI assigned a sample identification number (PTS Sample ID) to each sample using the following form:

R-MM-L-NNN

where R is a single-digit indicator of the testing round in which the sample was to be analyzed (R=1, 2 or 3),
MM was a two-letter indicator of the sample matrix and lead level (DL for dust/low, DM for dust/medium, PT for paint),
L was an unique identifier of the laboratory to which the batch was sent by the client (L=1, 2,...; L=1 if the client will be sending samples of the given matrix/lead level to only one laboratory), and
NNN was a three-digit sequential ID number that uniquely identifies each sample having common values for R and MM (example : 001-999).

The PTS Sample ID was copied onto standard self stick labels which were subsequently attached to the inside of a Ziplock™ bag. The labels were placed on the inside of the bags so that in the event the label fell off, it would still be associated with the sample bag. The double-blind dust and/or paint was transferred to the client-supplied containers and the containers placed into the pre-labeled plastic bags.

After assigning sample labels to the sample containers, RTI recorded the PTS Sample ID on a copy of the Sample Tracking and Analysis Report Form (Figure 1). The sample matrix of each sample type (dust or paint) , the weight of the sample placed into the sample container, and the date each sample was shipped to the clients was recorded. One copy of this form was made for retention at RTI, and the original was sent with the samples to the clients.

5.2 DISTRIBUTION

A total of 14 clients participated. The original mail-out of materials to 10 clients was on February 27, April 3, and June 1, 1998. Four clients joined the program late and received materials in April, May and June.

A copy of the Material Safety Data Sheet (MSDS), (Appendix A) for the lead containing materials was sent once to each client along with the instruction that it was for the clients use only and **not** to forward the MSDS with the double-blind samples to their laboratories.

NLLAP Double-Blind Proficiency Testing Pilot Program
Sample Tracking and Analysis Report Form

Client Information: (1)

Client:		
Address:		
City:	State:	Zip:
Telephone:	Fax:	
Responsible Party:		

Laboratory Information: (2)

Lab:		Lab ID#:
Address:		
City:	State:	Zip:
Telephone:	Fax:	
Responsible Party:		

Analysis Method: (17)

Paint: ☐ ICP ☐ GFAAS ☐ FAAS ☐ Other _______________
Dust: ☐ ICP ☐ GFAAS ☐ FAAS ☐ Other _______________

Section A: Sample Distribution from Proficiency-Testing Service (PTS) to the Client

PTS Sample ID (3)	Sample Matrix (4)	Sample Weight (grams) (5)	Date Shipped to Client (mm/dd/yy) (6)	Initials (7)	Date Received by Client (mm/dd/yy) (8)	Initials (9)
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					

Section B: Client Cross-Reference and Shipment to Laboratory

PTS Sample ID (10)	Client Sample ID (11)	Initials (12)	Date (mm/dd/yy) (13)	Date Shipped to Laboratory (mm/dd/yy) (14)	Initials (15)

Section C: Laboratory Results

Client Sample ID (16)	Reported Lead Concentration (18)	Units of Concen-tration (19)	Date Results Received by Client (mm/dd/yy) (20)	Initials (21)	Verification (22)	Date (mm/dd/yy) (23)

Numbers in parentheses in the column headings refer to citations in the *Standard Operating Procedure for Proficiency Sample Tracking and Data Reporting*.

SECTION 6.0
REFERENCES

1. Binstock, D.A., D.L. Hardison, P.M. Grohse, and W.F. Gutknecht, “Standard Operating Procedures for Lead in Paint by Hotplate- or -Microwave-based Acid Digestions and Atomic Absorption or Inductively Coupled Plasma Emission Spectrometry.” NTIS Publication N. PB 92-114172, EPA Contract No. 68-02-4550, September 1991.

APPENDIX A

MATERIAL SAFETY DATA SHEET

APPENDIX B

NLLAP Double-Blind Proficiency Testing Pilot Program

Script for Telephone Recruitment of Laboratory Clients

State: ________

Client name: ________________________________ Client ID #: ___________

Telephone number: __________________

Date of call: ________________ Time of call: _____________ AM PM

Caller: _________________________

Hello, this is (*state your name*). I understand that (you/your organization) send environmental samples to analytical laboratories to evaluate whether a home contains lead hazards. Who can I speak with who is responsible for overseeing the collection of dust or paint samples in homes and the shipping of these samples to a laboratory?

Obtain the contact name: ____

Obtain the title/position of contact: ____

Continue once the contact is on line.

Hello, my name is (*state your name*). I (work for/am a contractor to) the EPA's Office of Pollution Prevention and Toxics in Washington, DC. I understand that in the process of evaluating whether lead hazards are present in a home, you collect dust or paint chip samples from the home and send them to an analytical laboratory to determine the amount of lead in these samples.

Question #1: Is this true? ☐ No *Terminate the call*
☐ Yes *Continue*

Then you may be familiar with EPA's National Lead Laboratory Accreditation Program, or NLLAP. NLLAP recognizes analytical laboratories which have demonstrated and meet the minimum standards for analysis of lead in dust, soil, and paint chips. The EPA is currently investigating whether to implement a new procedure within the NLLAP which would improve the way by which laboratories are tested on their ability to analyze and report amounts of lead in environmental samples, and you may be able to help us in this study.

Question #2: Would you mind if I fill you in very briefly on this new procedure?

☐ No *Proceed* ☐ Yes *Terminate the call*

Currently in the NLLAP, approximately every three months, a proficiency testing service sends a special batch of proficiency-test samples to a laboratory, then the laboratory analyzes these samples and reports back the results. The proficiency testing service knows how much lead is in each sample, so they can compare what the laboratory reports with what is actually in the sample. However, one disadvantage to this procedure is that the laboratory knows that they are being evaluated based on the results they report for these special samples. The evaluation would be better if the laboratory did not know when they were analyzing these proficiency-test samples.

The new procedure which EPA is considering would have the proficiency testing service send the proficiency-test samples to clients of the laboratories, rather than directly to the laboratories. The client would place these samples within batches of their own field samples and send the batches off to the laboratory for analysis. The proficiency-test samples would remain anonymous within the batches, and so the laboratory would not know that they are testing proficiency-test samples. This approach is called a double-blind approach and leads to a more accurate evaluation of a laboratory's routine performance in analyzing samples that they receive from their clients.

Here's where you come in. We need to identify clients of NLLAP-accredited laboratories who would be willing to participate in a pilot study that will evaluate the feasibility of developing a double-blind program. Each client in this study will receive no more than 18 proficiency-test samples (perhaps less) free of charge around the end of February. They would be given explicit instructions on how to incorporate these samples into three batches of their field samples, submit these batches to the laboratory, and report the results of the proficiency-test samples back to us. These clients would be involved in the study for about three months, and it would take a minimal amount of effort on their part to participate.

Question #3: Does this sound like a study that you may be interested in participating in?

☐ Yes *Proceed* ☐ No *Terminate the call*

In order to determine which role, if any, you may be able to play in this pilot study, I need to ask you a few questions about the types and numbers of field samples you submit to a laboratory for analysis.

Question #4: First, do you send either dust-wipe or paint chip samples to an NLLAP-accredited laboratory for analysis? ☐ No *Terminate the call*
☐ Yes *Proceed*

Question #5a: Do you currently run a double-blind test of your own on an NLLAP-accredited laboratory? ☐ No *Go to Question #6a*
☐ Yes *Proceed*

Question #5b: In the double-blind test, do you submit blank samples to the laboratory, or do you spike samples with known amount of lead prior to sending the samples to the laboratory?

☐ Submit blank samples *Go to Question #6a*
☐ Submit spiked samples, or don't know *Proceed*

Question #5c: Are there any NLLAP-accredited laboratories that you contract with in which you do not run double-blind tests? ☐ No *Terminate the call*
☐ Yes *Proceed*

In the remaining questions, we are interested in only NLLAP-accredited laboratories that you send samples to for which you do not run double-blind tests.

Dust-Wipe Samples

Question #6a: Do you send at least one batch of dust-wipe samples every month to an NLLAP-accredited laboratory? ☐ No *Proceed to Question 7a* ☐ Yes

Question #6b: Approximately how many dust-wipe samples do you place in a typical batch? _____________

Question #6c: Can you tell me the type or brand of dust wipe you use in the field? *If they don't know, ask if it is a hand-towelette versus a baby wipe.*

____

Question #6d: What type and brand of sample container do you place the dust-wipe sample in when sending the sample to the laboratory?

____

Question #6e: In general, do you use only one dust-wipe to collect a dust sample, or do you use more than one wipe to collect a single sample?

☐ Use only one wipe per sample ☐ Use multiple wipes per sample

Paint Chip Samples

Question #7a: Do you send at least one batch of paint-chip samples every month to an NLLAP-accredited laboratory? ☐ No *If Question #6a was yes, then skip Questions #7b and #7c*
Otherwise, terminate the call
☐ Yes

Question #7b: Approximately how many paint-chip samples do you place in a typical batch? ____________

Question #7c: What type and brand of sample container do you place the paint chip sample in when sending the sample to the laboratory?

____

Question #7d: Do you prepare the paint-chip sample in any way prior to placing them in sample containers for shipment to the laboratory, such as grinding them to a powder?

☐ No ☐ Yes _____________________________________

Laboratory

Question #8a: What are the names and locations of NLLAP-accredited laboratories that you send dust-wipe and/or paint chip samples to on at least a monthly basis? *Note that these should not be laboratories that the client is currently sending double-blind samples to already.*

Lab #1: ____

Lab #2: ____

Lab #3: ____

If the answers to Question #6a and #7a are both yes, then ask Question #8b. Otherwise, skip to Question #8d.

Question #8b: Do you submit both dust-wipe and paint chip samples to the same laboratory?

☐ No *Skip to Question #8d* ☐ Yes *Proceed*

Question #8c: To which labs do you send dust-wipe samples, and to which do you send paint chip samples? *Respond according to responses to Questions #6a and #7a.*

Dust-wipe samples: ☐ Lab #1 ☐ Lab #2 ☐ Lab #3
Paint-chip samples: ☐ Lab #1 ☐ Lab #2 ☐ Lab #3

Question #8d: Do you know approximately how many samples the laboratory tests at a given time, like per month? *Respond according to responses to Questions #6a and #7a. We want to get a basic idea on whether the lab is large or small.*

Lab #1: Dust-wipe______________________Paint-chip_____________________

Lab #2: Dust-wipe______________________Paint-chip_____________________

Lab #3: Dust-wipe______________________Paint-chip_____________________

Question #8e: Do you know the analytical method used by the laboratory to analyze the samples? *Respond according to responses to Questions #6a and #7a.*

Lab #1: Dust-wipe______________________Paint-chip_____________________

Lab #2: Dust-wipe______________________Paint-chip_____________________

Lab #3: Dust-wipe______________________Paint-chip_____________________

Question #8f: Do you know the analytical method's detection limit for lead? *Respond according to responses to Questions #6a and #7a.*

Lab #1: Dust-wipe______________________Paint-chip_____________________

Lab #2: Dust-wipe______________________Paint-chip_____________________

Lab #3: Dust-wipe______________________Paint-chip_____________________

Those who participate in the pilot study will be required to submit as many as 30 empty, unused sample containers to the proficiency testing service in which they will place the proficiency-test samples.

Question #9: Would you agree to do this if you were involved in this study?

☐ Yes ☐ No

Thank you very much for your time and for your interest in this pilot study. We will get back to you by the end of the month on whether or not you will be selected to participate in this study. Could I get your address?

Street Address: ____

City: ___________________________ State: ________ Zip Code: __________

Verify name of contact.

APPENDIX C

January 26, 1998

(Name of Contact)
(Name of Business)
(Address)
(City, ST ZIP)

Dear (Name):

Thank you for expressing an interest in participating in a double-blind proficiency-testing pilot program being conducted by the U.S. Environmental Protection Agency. **We are happy to inform you that you have been selected to participate in this program!**

As we discussed with you on the telephone, the primary objective of this pilot is to evaluate the feasibility of implementing a double-blind program as a supplement to the current single-blind laboratory proficiency-testing program within the National Lead Laboratory Accreditation Program (NLLAP). In a double-blind program, the proficiency-testing service would supply proficiency-test samples to the clients of laboratories, rather than directly to laboratories as a single-blind program does. The clients would then incorporate these samples within batches of field samples, submit the batches to the laboratories, and report the analytical results of the proficiency-test samples back to the proficiency-testing service. Therefore, unlike a single-blind program, a double-blind program does not allow the laboratories to know when they are analyzing proficiency-test samples, thereby allowing their routine performance to be more accurately measured.

The attachment to this letter provides instructions for you to follow as a participant in this pilot program. Through the course of the program, **we will be providing you with a total of (specify: #) dust-wipe samples and (specify: #) paint chip samples for you to submit to (specify: name of laboratory here) for analysis.** We will provide you with these samples over three testing rounds: in March, April, and June, 1998. Within each testing round, we will ask you to incorporate the samples you receive into the next available batch of field samples that you will submit to **(specify: this laboratory, these laboratories)** for analysis.

In order to ensure that the laboratory cannot discern the proficiency-test samples from the field samples, we need to place the proficiency-test samples in the same containers that you will use for the field samples. Therefore, **please submit (specify: #) empty, unused sample containers by February 13, 1998,** to Ms. Laura Hodson at the address specified on the attachment, so that we can place our samples in the same containers you will be using.

During the pilot program, please regard all aspects of the program as **confidential and program-sensitive**. You will be able to find out the results of the analysis of these proficiency-test samples at the end of the pilot program, upon release of the program's final report.

We will be contacting you by telephone in the next week to review the instructions in this letter and on the attached sheet and to verify your participation in the program. We look forward to talking with you again and answering any questions that you may have. Meanwhile, please call Bob Lordo of Battelle at (614) 424-4516 if you have any questions on your acceptance in the program. Thank you again for your participation!

Sincerely,

John Scalera
Office of Pollution Prevention and Toxics

NLLAP Double-Blind Proficiency-Testing Pilot Program

Information and Instructions to Participating Laboratory Clients
REVISED

Thank you for agreeing to participate in this double-blind proficiency-testing pilot program! The information that we can gather in this program will be very useful in developing a double-blind program within the NLLAP. Such a program will eliminate the need for laboratory clients to perform their own double-blind procedures, which will greatly benefit all lead inspectors and risk assessors who submit environmental samples to laboratories for lead analysis.

During the pilot program, you will interact with the program's proficiency-testing service who will manage the shipping of proficiency-test samples and the collecting of analysis results on these samples. Your **primary contact at the proficiency-testing service** on any questions you may have, issues that arise, and quality assurance issues to report, is:

Fred Grunder, CIH
Manager, Laboratory Accreditation Programs
American Industrial Hygiene Association (AIHA)
2700 Prosperity Avenue, Suite 250
Fairfax, VA 22031
phone: 703/849-8888
fax: 703/207-3561
e-mail: fgrunder@aiha.org

If you are unable to reach the primary contact, you may contact Mr. Carl Bell at AIHA (at the same address and telephone numbers as above), or the NLLAP staff at the U.S. Environmental Protection Agency at 202/260-6709.

The proficiency-testing service will provide you with proficiency-test samples in each of **three testing rounds: in March, April, and June, 1998**. This attachment provides you with information such as how to place the proficiency-test samples for a given testing round within your next available batch of field samples, how to properly track these samples while in your possession, and how to report the results of these samples to the proficiency-testing service.

Although we are unable to reimburse you for shipping/postage costs and for costs to supply the proficiency-test service with sample containers, we will be happy to reimburse you for the cost of laboratory analysis of the proficiency-test samples we supply to you in this program. To receive reimbursement of the analysis costs, please provide a copy of the laboratory's invoice, showing either 1) the cost per sample, or 2) the number of samples analyzed and the total analysis cost, to:

Robert Lordo, Ph.D.
Battelle

505 King Avenue
Columbus, OH 43201

You can also fax this information to Dr. Lordo at 614/424-4516.

In order to preserve the double-blind nature of this pilot program, while ensuring proper sample tracking and data reporting, we have identified **seven primary tasks** for you to follow in the program. Each task will require only a minimal effort on your part. You need perform Task 1 only once; Tasks 2 through 6 will be accomplished three times, once in each testing round.

Task 1: Please **provide clean, unused sample containers** in which the proficiency-testing service will place your proficiency-test samples. The number of containers to submit is specified in the letter accompanying these instructions. Please submit these containers via priority mail or first-class mail **by February 13, 1998,** to the following address:

Laura Hodson, CIH
Center for Environmental Measurements and Quality Assurance
Research Triangle Institute
Research Triangle Park, NC 27709

Within your shipment of the sample containers, please **notify us on the proper address that we should use in submitting the packages of proficiency-test samples to you.**

Task 2: Each shipment of proficiency-test samples that you will receive will be accompanied by the attached *Sample Tracking and Analysis Report Form*, with one copy included for every laboratory to whom you will be shipping the samples for analysis. This form is used to properly track the proficiency-test samples through the program and to report the analytical results of these samples.

C. When you receive the proficiency-test samples, please store them in a **locked area with limited access** (e.g., cabinet, closet) until you place them within your next available batch of field samples.

D. **Do not open** or otherwise tamper with the sample containers or their contents.

E. Please store the accompanying *Sample Tracking and Analysis Report Forms* in a **locked area** when not in use.

Task 3: For the *Sample Tracking and Analysis Report Forms* that you receive,

A. Please **review the information which the proficiency-testing service provides in the block titled "Client Information (1)"** and make any necessary changes. (Note that the proficiency-testing service will also provide sample information in columns (3) through (7) on the form.)

B. In the upper right of the form, within the block labeled "Laboratory Information (2)", please **supply information on the specified laboratory** that will receive the proficiency-test samples for analysis.

C. Please **specify the date you received the proficiency-test samples** in the column of the form labeled "Date Received by Client (8)" and **place your initials** (indicating that this date was recorded correctly) in the column labeled "Initials (9)".

Task 4: Upon your receipt of proficiency-test samples in a given testing round, please use the *Sample Tracking and Analysis Report Forms* to identify which samples are to go to which laboratories for analysis (if you are to submit samples to multiple laboratories). Then, place a laboratory's proficiency-test samples **randomly** within your next available batch of field samples that are earmarked for analysis at that laboratory. To ensure that the proficiency-test samples remain properly identified through the program, please perform the following **while treating the proficiency-test samples no differently from the field samples:**

A. Please **assign identifications (IDs) to all samples in the batch**.

B. For each proficiency-test sample, please **record the PTS Sample ID** (i.e., the ID specified on the label when you receive the samples) in the column labeled "PTS Sample ID (10)" within Section B of the *Sample Tracking and Analysis Report Form*, **and the ID which you assign to the sample** in the column labeled "Client Sample ID (11)." When recording these two IDs, please ensure that the two IDs on a given row of the table are for the **same sample**.

C. For each proficiency-test sample, please **remove the sample container from the outer plastic bag, noting the PTS Sample ID that is on this outer bag, then place a label on the sample container containing the ID you assign to the sample.**

D. Once you have verified that each proficiency-test sample is properly identified in the batch and on the *Sample Tracking and Analysis Report Form*, please **initial and date** the columns labeled "Initials (12)" and "Date (13)".

Task 5: Please perform the following when shipping a batch of field and proficiency-test samples to the laboratory for analysis:

A. Please **include any forms you would routinely send with the batch**, such as your chain-of-custody form. **(Do not send the *Sample Tracking and Analysis Report Form* with the samples!)** If your form includes such information as sample area, location, and substrate for each sample, please specify this information for the proficiency-test samples (even if you have to make up the information) so that the laboratory believes that these are actual field samples. For example, you may want to specify that a dust-wipe proficiency-test sample was "collected" from a one square-foot area on the living room floor. If you are receiving paint-chip proficiency-test samples, please refer to these samples as "**paint chips from brick-**

or concrete-surfaces" or "**dust contaminated with paint**" due to their small particle sizes.

B. Please **specify the date shipped to the laboratory** in the column labeled "Date Shipped to Laboratory (14)" on the *Sample Tracking and Analysis Report Form* and **place your initials** in the column labeled "Initials (15)" on the form.

C. On the day that you ship the samples, please **fax a copy of the *Sample Tracking and Analysis Report Form* (completed through Section B) to Fred Grunder of AIHA** (fax number: 703/207-3561).

D. A successful pilot program will provide important information on double-blind laboratory accreditation, which will ultimately benefit you and others who employ laboratories for accurate analysis of lead in environmental samples. Therefore, **please do not discuss the proficiency-test samples with the laboratory in any way that would indicate that they are not routine samples, and do not divulge to the laboratory your participation in the pilot program.**

Task 6: Once the laboratory has provided you the analytical results, please perform the following for each proficiency-test sample in the batch:

A. Please **record the sample ID that you assigned to each proficiency-test sample** in the column labeled "Client Sample ID (16)" in Section C of the *Sample Tracking and Analysis Report Form*, then **record the date you received the results, along with your initials**, in the columns labeled "Date Results Received by Client (20)" and "Initials (21)" on the form.

B. Please **specify the analysis method used by the laboratory** in the box labeled "Analysis Method (17)" on the *Sample Tracking and Analysis Report Form* (ICP=inductively coupled plasma-atomic emission spectroscopy, GFAAS= graphite furnace atomic absorption spectroscopy, FAAS=flame atomic absorption spectroscopy).

C. Please **specify the lead concentration reported by the laboratory for the proficiency-test sample, as well as the units of measurement,** in the columns labeled "Reported Lead Concentration (18)" and "Units of Concentration (19)" in Section C of the *Sample Tracking and Analysis Report Form*.

D. Please review all entries on the form for any transcription errors, then **initial and date** the column of the *Sample Tracking and Analysis Report Form* labeled "Verification (22)" and "Date (23)".

E. Please **send a copy of the final completed form to Fred Grunder of AIHA via fax and mail within three working days after receipt of analysis results from the laboratory** (fax number: 703/207-3561; mailing address given on the first page of this attachment).

Task 7: Throughout the pilot program, please feel free to **report any quality assurance issues** (e.g., compromised proficiency-test samples, errors in tracking or reporting) to **Fred**

Grunder or Carl Bell of AIHA (phone number: 703/849-8888; fax number: 703/207-3561; mailing address given on the first page of this attachment).

The *Sample Tracking and Analysis Report Form* is attached for your reference.
Thank you again for your participation!

NLLAP Double-Blind Proficiency Testing Pilot Program
Sample Tracking and Analysis Report Form

Client Information: (1)

Client:		
Address:		
City:	State:	Zip:
Telephone:	Fax:	
Responsible Party:		

Laboratory Information: (2)

Lab:		Lab ID#:
Address:		
City:	State:	Zip:
Telephone:	Fax:	
Responsible Party:		

Analysis Method: (17)

Paint: ☐ ICP ☐ GFAAS ☐ FAAS ☐ Other _______________
Dust: ☐ ICP ☐ GFAAS ☐ FAAS ☐ Other _______________

Section A: Sample Distribution from Proficiency-Testing Service (PTS) to the Client

PTS Sample ID (3)	Sample Matrix (4)	Sample Weight (grams) (5)	Date Shipped to Client (mm/dd/yy) (6)	Initials (7)	Date Received by Client (mm/dd/yy) (8)	Initials (9)
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					
	☐ Paint ☐ Dust Wipe					

Section B: Client Cross-Reference and Shipment to Laboratory

PTS Sample ID (10)	Client Sample ID (11)	Initials (12)	Date (mm/dd/yy) (13)	Date Shipped to Laboratory (mm/dd/yy) (14)	Initials (15)

Section C: Laboratory Results

Client Sample ID (16)	**Reported Lead Concentration (18)**	**Units of Concen-tration (19)**	**Date Results Received by Client (mm/dd/yy) (20)**	**Initials (21)**	**Verification (22)**	**Date (mm/dd/yy) (23)**

Numbers in parentheses in the column headings refer to citations in the *Standard Operating Procedure for Proficiency Sample Tracking and Data Reporting*.