Evaluation of Filter Paper Blood Lead Methods: Results of a Pilot Proficiency Testing Program

Noel V. Stanton,1* Jane M. Maney,1 and Robert Jones2

Background: Lead testing on dried filter paper (FP) blood spots is used routinely by some laboratories for lead poisoning screening. Proficiency testing (PT) as required under CLIA ‘88 laboratory regulations has not been available for these methods.

Methods: We describe a suitable PT scheme and evaluate FP laboratory performance based on program results. Monthly testing events consisting of five FP specimens were provided to six participating laboratories. Results were evaluated against target values determined by referee laboratories.

Results: Preliminary FP laboratory results showed poor agreement with specimen target values, exhibiting a mean absolute bias of 0.29 μmol/L (5.9 μg/dL). Five of six participating laboratories demonstrated significant improvement in later testing events, with bias decreasing to 0.12 μmol/L (2.5 μg/dL). Performance varied widely between the participating laboratories and appeared to be method dependent. When evaluated using CLIA blood lead acceptability criteria, the proportion of acceptable individual specimen results (n = 35) ranged from 54% to 100%. On a testing event basis (n = 7), the proportion of acceptable events ranged from 29% to 100%.

Conclusions: A suitable FP PT program now exists to capably assist and monitor FP laboratories. Based on overt PT results, properly utilized FP testing methods can accurately measure blood lead concentration.

Pediatric lead poisoning is a prevalent environmental health problem, with ~890,000 US children believed to have blood lead concentrations exceeding the current acceptable threshold of 0.48 μmol/L (10 μg/dL) (1). Because most cases of lead poisoning are asymptomatic, laboratory measurement of blood lead concentration is required for proper diagnosis. Formerly, measurement of erythrocyte protoporphyrin was utilized as the primary screening test for childhood lead poisoning detection. However in 1991, recommendations were issued by the CDC that called for the universal screening of young children using direct blood lead measurement (2). These recommendations led to greatly increased demand for blood lead testing and heightened interest in novel, rapid, and inexpensive sampling and analysis techniques.

The use of blood spots dried on filter paper (FP) has been investigated as an alternative to liquid blood for lead measurement. Possible advantages of FP sampling include greater ease of collection, simpler packaging, lower shipment cost, and long-term specimen stability. Methods utilizing FP samples were first described in the 1970s as a modification of the Delves cup flame atomic absorption (AA) method for liquid blood lead testing and heightened interest in novel, rapid, and inexpensive sampling and analysis techniques.

Recent publications have described several new FP-sample testing methods that use a variety of analytic approaches. Methods using electrothermal atomization (graphite furnace) AA measurement after sample extraction with solutions of HNO3 or (NH4)2HPO4 (12), inductively coupled plasma mass spectrometry measurement after HNO3 extraction with yttrium internal standard (13), and modifications of the original Delves cup AA procedure incorporating a preashing step (14, 15) have all been reported since 1991. These recent publications largely conclude that FP collection is a viable alternative for blood lead screening purposes, and a few laboratories now routinely perform FP blood lead testing on pediatric specimens.

The reliability of FP methods has been questioned (16–19), and continues to generate discussion (20–22).

1 Wisconsin State Laboratory of Hygiene, Toxicology Section, 2601 Agriculture Dr., P.O. Box 7996, Madison, WI 53707-7996.
2 Centers for Disease Control and Prevention, National Center for Environmental Health, Environmental Health Lab Services, 4770 Buford Hwy., Atlanta, GA 30341.
*Author for correspondence. Fax 608-224-6259; e-mail nvstox@mail wlsl.wisc.edu.
Received April 16, 1999; accepted August 21, 1999.

Nonstandard abbreviations: FP, filter paper; AA, atomic absorption; PT, proficiency testing; WSLH, Wisconsin State Laboratory of Hygiene.

Based in part on the data presented here, CDC has revised its position regarding the use of FP-based methods by CDC childhood lead poisoning grantees. This new position is described in a letter to those grantees dated February 25, 1999.
One major concern has been the absence of FP-based proficiency testing (PT) as required under the CLIA ‘88 laboratory regulations (23) and the associated external method validation data available through PT participation. Enlisting the cooperation of laboratories with FP testing expertise, we sought to establish a PT program that would be viable for all FP methods in current use. This report describes the characteristics of this pilot PT program and examines the results from the participating FP laboratories.

Materials and Methods

PROGRAM CONFIGURATION AND LABORATORY RECRUITMENT

The Wisconsin State Laboratory of Hygiene (WSLH) has administered the grant-supported (Health Resources and Services Administration, US Department of Health and Human Services) National Blood Lead PT Program since 1988. This established program, which has been described elsewhere (24), provided an initial model for the development of a PT program for FP-based blood lead methods. Laboratories with experience in FP testing were identified through announcements to existing PT program participants, information and announcements provided by the CDC, and through the review of recent FP literature. These efforts identified six laboratories (five in the US and one in Canada) willing to participate. Based on the FP literature, the extant PT program, and input from the participating laboratories, a draft program framework was developed and distributed for review. Representatives from WSLH and five of the participants then met in July 1997 to finalize program characteristics. The remaining participant reviewed a summary of the meeting. The protocols that follow reflect the consensus of WSLH and all of the participating laboratories.

BLOOD DONORS AND COLLECTION

Fresh human blood with physiologically bound lead was used for the FP specimen matrix. Blood that was hemeolyzed or to which a lead solution had been added were reported to be unsuitable because of FP blood spot characteristics that differ from fresh human blood (10). Nonhuman blood was considered unsuitable for the same reason. All procedures were approved by the University of Wisconsin-Madison Center for Health Sciences Human Subjects Committee and the CDC Internal Review Board, and are in accordance with the Helsinki Declaration. Volunteer donors were recruited from a cohort of mixed sex and race industrial workers with occupational lead exposure. The cohort was monitored independently for their lead exposure in compliance with Occupational Safety and Health Administration guidelines. Six nonexposed adults were also recruited to provide specimens with background lead concentrations.

After giving informed consent, each donor underwent a standard venipuncture by trained medical personnel, using equipment obtained from Becton Dickinson. Three 10-mL blood specimens were collected in tubes containing liquid EDTA anticoagulant. All blood collection containers used in the program were lot tested for the presence of lead by a WSLH adaptation of published protocols (20). Each volunteer donation was a discrete FP PT specimen. Blood from 45 donors was used in the study. The donors were questioned about medication use in the 24-h period before collection, and a “Yes” or “No” response was recorded. The blood specimens were then transported and stored at refrigerated temperatures.

SPECIMEN PREPARATION

To minimize the possibility of environmental contamination, specimen preparation steps were performed in a trace metal clean room conservatively rated as Class 1000 (M4.5). The following steps were performed within 30 h of blood collection. Contents of the three tubes from an individual donor were combined and gently mixed in a 50- to 100-mL beaker by a magnetic stirring device. Blood was then spotted onto Schleicher & Schuell no. 903 filter paper sheets, which were suspended horizontally in air by clips. Five of the six participating laboratories received FP sheets containing three 50-μL spots. These sheets were spotted using an Eppendorf positive displacement repeating pipette with disposable tip assemblies. Laboratory A requested sheets with five, and later six, spots of 30 μL each to comply with their unique FP specimen submission requirements. This was done with an air displacement 30-μL Eppendorf pipette with disposable tips. The accuracy of pipettes used for FP spotting was verified before use. The remainder of the blood was dispensed in 2-mL aliquots into 3-mL additive-free evacuated blood tubes, using an Eppendorf positive displacement repeating pipette. These aliquots were then provided to referee laboratories for value assignment purposes.

After spotting, the FP sheets remained suspended horizontally to allow the blood spots to air dry. The sheets were dried for an average of 75 min (range, 58–103 min) and then returned to zip-lock plastic bags for storage at room temperature. Subsequent distribution of the FP specimens to participants took place within 60 days. The aliquots of liquid blood were stored under refrigeration until distribution to reference laboratories for value assignment. These specimens were distributed within 1 week.

The FP sheets had designated handling and spotting areas marked on them. Other than the zip-lock bags the sheets were stored in, the spotting areas of the sheets did not come in contact with any surfaces during the spotting or drying process. Specimens from the first three testing events (October through December 1997) were spotted on blank FP sheets that had been provided by the individual participating laboratories. Although the sheets varied considerably in size and appearance, they shared the common characteristics of Schleicher & Schuell no. 903 composition, designated handling and spotting areas, and prior suitability testing by the participants. All other specimens were spotted on custom Schleicher & Schuell no. 903 sheets obtained for
the program and designed to specifications suitable for the participants. Before use for PT purposes, these sheets were also evaluated for suitability by all of the participating laboratories.

**BLOOD SPECIMEN CHARACTERIZATION**

The blood specimens were visually inspected for hemolysis and lipemia immediately before spotting on the FP sheets. Specimens exhibiting hemolysis were rejected for study purposes. The presence of lipemia was noted but was not a criterion for specimen exclusion. The hemoglobin concentration was measured for all specimens, using an Instrumentation Laboratory Model 682 Co-oximeter. The homogeneity of blood during the spotting procedure was verified by measurement of hemoglobin concentrations on serial aliquots of a representative blood specimen. Aliquots dispensed at 15-s intervals for 6 min gave the following results: mean ± SD, 2279 ± 25 μmol/L (14.7 ± 0.16 g/dL); range, 2233–2310 μmol/L (14.4–14.9 g/dL).

No time-dependent trends were observed.

Blood lead concentrations were determined using the protocol of the established PT program (24). Briefly, 10 referee laboratories were provided 2-mL aliquots of liquid blood for each specimen, which were analyzed using their routine procedures. Methods included electrothermal atomization AA and anodic stripping voltammetry techniques. Target values were then determined by the mean of the referee laboratory results, after deletion of outliers. Results falling outside the range, target value ± 0.19 μmol/L (4 μg/dL) or 10%, were excluded as outliers.

**REPORTING AND EVALUATION**

Monthly FP PT events commenced in October 1997. Each monthly testing event consisted of five specimens. The FP specimens, along with a report form, were shipped by surface mail in Tyvek® envelopes. Participants then re-specified specimens, along with a report form, were shipped by monthly testing event consisted of five specimens. The FP event score was determined by the mean of the referee laboratory results, after deletion of outliers. Results falling outside the range, target value ± 0.19 μmol/L (4 μg/dL) or 10%, were excluded as outliers.

**REFERENCE LABORATORY RESULTS**

Target values for the October-December 1997 FP PT events were determined by the mean of nine referee laboratory results. Results from one of the referee laboratories were excluded from target value calculations for these specimens because of a single outlying result. This policy was consistent with that of the established blood lead PT program. Target values for all other events (specimen n = 30) were the mean of 10 referee laboratory results. The lead concentration of the specimens ranged from 0.05 to 1.83 μmol/L (1–38 μg/dL), with a range of 0.53–2.23 μg/dL; the values generally increasing as a function of concentration. This level of uncertainty is comparable to that observed in the established PT program.

**PARTICIPANT GROUP RESULTS**

The individual participant results for both preliminary and evaluated specimens, plotted as a function of deviation from the target value, are shown in Fig. 1. Results on the preliminary challenges, to the left of the vertical line in Fig. 1, exhibit wide scatter. Absolute deviation from the target value for this group of specimens (mean ± SD), was 0.29 ± 0.13 μmol/L (5.9 ± 2.7 μg/dL) and ranged from +0.77 μmol/L (+16 μg/dL) to −0.92 μmol/L (−19 μg/dL). Although the observed data scatter for the group as a whole was bidirectional, the bias observed in results from individual laboratories, when present, was unidirectional.

Following the preliminary events, all specimens were evaluated using the CLIA acceptability criteria. Shown to the right of the vertical line in Fig. 1, a substantial
improvement in participant accuracy coincided with the initiation of PT evaluation. The mean absolute deviation from target values decreased by 50%, to 0.12 ± 0.05 mmol/L (2.5 ± 1.0 mg/dL), with a range from 1.03 mmol/L (11.1 mg/dL) to 2.30 mmol/L (21.5 mg/dL). Five of the six individual laboratories exhibited statistically significant improvement; laboratory A, P = 0.012; laboratories B, C, and D, P < 0.001; and laboratory E, P = 0.004. The performance of laboratory F did not change significantly between the preliminary and evaluated specimens. The improvements can be attributed to changes made by participants, although the specific nature of these changes is not known. No changes to program procedures were made.

The 35 FP PT challenges evaluated by the six participants produced 210 discrete results. Of these, 176 (84%) fell within the CLIA acceptable range. Of the 34 results falling outside the acceptable range, 29 (85%) fell below the minimum threshold, whereas only 5 (15%) were unacceptably high.

METHOD-SPECIFIC PERFORMANCE

The six participating FP laboratories used three distinct methodologies. Laboratories A, E, and F extracted blood spots with an (NH₄)₂HPO₄-Triton X-100 solution, followed by electrothermal atomization AA measurement. These laboratories utilized aqueous calibrators, based on the method described by Yee and Holtrop (10). Grouped by these methodologies, regression of the individual laboratory results against target values is shown in Fig. 2. Data from the participating laboratories correlated well with target values. Laboratories A, E, and F demonstrated the highest correlation, r ≥ 0.98. Regression slopes for laboratories E and F were not statistically different from unity, whereas laboratory A showed a small but significant (P < 0.05) negative bias. Laboratories B, C, and D exhibited substantial negative bias in the slope of their regressions, approaching 25% in the case of laboratory C. The bias was highly significant statistically (P < 0.001) for each of these laboratories. In addition, the y-intercepts of laboratories B and D differed significantly from zero (P < 0.01 for both), exceeding 0.13 mmol/L (2.7 μg/dL) in both laboratories.

Fig. 1. Individual FP laboratory results, plotted as the deviation from target values (n = 45). Dates of the five specimen PT events are shown on the x-axis. The tick marks denote individual PT specimens. The vertical line separates ungraded preliminary PT specimens from evaluated specimens. The dotted horizontal lines are CLIA acceptability limits. 1 μg/dL = 0.4826 μmol/L.

Fig. 2. Linear regression analysis of FP laboratory results vs target values, grouped by similar FP measurement methods (n = 35). 1 μg/dL = 0.4826 μmol/L.
Table 1. Predicted FP results at selected CDC blood lead decision levels.

<table>
<thead>
<tr>
<th>Laboratory</th>
<th>Threshold value, μmol/L (µg/dL)</th>
<th>0.48 (10)*</th>
<th>0.97 (20)*</th>
<th>2.17 (45)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.44 (9.2)</td>
<td>0.90 (18.6)</td>
<td>2.04 (42.2)</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.53 (11.0)</td>
<td>0.93 (19.3)</td>
<td>1.93 (39.9)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.38 (7.8)</td>
<td>0.74 (15.3)</td>
<td>1.65 (34.1)</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>0.51 (10.5)</td>
<td>0.88 (18.3)</td>
<td>1.83 (37.8)</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>0.48 (9.9)</td>
<td>1.00 (20.6)</td>
<td>2.29 (47.3)</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0.51 (10.6)</td>
<td>1.00 (20.7)</td>
<td>2.21 (45.8)</td>
<td></td>
</tr>
</tbody>
</table>

* CDC decision level.

**PREDICTED RESULTS AT CDC DECISION LEVELS**

FP results at significant blood lead decision level concentrations were predicted using the participant regression equations. Predicted FP results at CDC decision level lead values of 0.48, 0.97, and 2.17 μmol/L (10, 20, and 45 µg/dL) (2) can be seen in Table 1. Because the highest PT specimen concentration was 1.83 μmol/L (38 µg/dL), the predicted results at the 2.17 μmol/L (45 µg/dL) decision level represent an extrapolation of the actual data. Predicted results fell within ± 0.05 μmol/L (1.0 µg/dL) for five of the participants at the 0.48 μmol/L (10 µg/dL) decision level and for three of the participants at the 0.97 μmol/L (20 µg/dL) decision level. The largest predicted differences at these decision levels were – 0.11 and – 0.23 μmol/L (– 2.2 and – 4.7 µg/dL), respectively. At the 2.17 μmol/L (45 µg/dL) decision level, predicted results from three laboratories fell within ± 0.14 μmol/L (2.8 µg/dL), but results from the other three laboratories had a predicted negative bias exceeding 0.24 μmol/L (5 µg/dL).

An evaluation of participant performance under the CLIA ‘88 acceptability criteria for blood lead is shown in Table 2. The CLIA regulations established performance criteria at several levels (23). Acceptable results for individual PT specimens are defined as the target value ± 0.19 μmol/L (4 µg/dL) or 10%, whichever is greater. CLIA mandates that testing events consist of five PT specimens. An event score of ≥80% (four of five) acceptable specimen results constitutes satisfactory event performance. Event scores, in turn, are evaluated on a cumulative basis. Satisfactory results on ≥67% (two of three) consecutive testing events constitute successful performance. The 35 evaluated PT specimens were provided as 7 five-specimen testing events, conforming to the CLIA configuration. The events took place on a monthly basis to accelerate data accumulation, rather than three times per year as specified under CLIA. Performance varied considerably among the participants. On an individual specimen basis, acceptable results ranged from 100% (35 of 35) attained by laboratories A and F to 54% (19 of 35) scored by laboratory C.

This scoring included seven alphanumeric results. Laboratory A did not quantify results exceeding 0.97 μmol/L (20 µg/dL) in the first evaluated testing event, reporting results as >0.97 μmol/L (>20 µg/dL) for three specimens with target values of 1.06, 1.16, and 1.83 μmol/L (22, 24, and 38 µg/dL). Laboratory D reported a result of >1.69 μmol/L (>35 µg/dL) for the specimen with a target value of 1.83 μmol/L (38 µg/dL). Laboratory A also reported results of <0.10 μmol/L (2 µg/dL) on specimens with target values of 0.05 and 0.10 μmol/L (1 and 2 µg/dL). Laboratory E reported a result of <0.05 μmol/L (<1 µg/dL) on a specimen with a target value of 0.14 μmol/L (3 µg/dL). Identical data points and the use of alphanumeric results led to fewer than 35 distinct data points on the plots in Fig. 2. All seven of the alphanumeric results were scored as acceptable for evaluation purposes.

At the testing event level, three of the six participants, laboratories A, E, and F, achieved satisfactory performance on 100% (seven of seven) of the testing events. The poorest performance was observed in laboratory C, which achieved satisfactory performance in 29% (two of seven) of the events. When the cumulative scoring criteria were used, laboratories A, E, and F also demonstrated successful performance throughout the seven testing events. With the exception of laboratory C, all of the laboratories were demonstrating successful performance at the completion of the final testing event.

**Discussion**

The FP PT scheme, developed by consensus of the participating laboratories, was a useful tool for assessing the performance of FP blood lead methods. Three distinct assessment methods, deviation from the target value, linear regression analysis, and evaluation using CLIA acceptability criteria, all yielded similar conclusions re-
garding performance of the six participating laboratories. In addition, the improvement in laboratory performance observed from the initial ungraded events illustrates the value of external PT in raising the quality of FP blood lead tests.

Some statements about individual laboratory performance can be made. Performance varied widely between individual laboratories. The best performance was exhibited by laboratories A, E, and F, all of which use an extraction reagent containing (NH₄)₂HPO₄ and Triton X-100 as well as blood-based calibrators. In contrast, the poorest performance, present as a negative bias, was demonstrated by laboratories C and D, both of which use a HNO₃/Triton X-100 extraction reagent and aqueous calibrators. The source of this negative bias is not well understood. Possible causes include matrix effects related to the use of aqueous calibrators or adverse effects of EDTA on the efficiency of the HNO₃ extraction. The EDTA present in the PT specimen blood spots would not be present in patient specimens submitted to these laboratories for routine analysis. In November 1997, laboratory D submitted a small (four specimens) experimental data set utilizing an alternative calibration of EDTA-anticoagulated blood rather than the aqueous calibration the laboratory routinely uses. This data set showed much closer agreement to target values, which is consistent with either of the proposed explanations. Unfortunately, laboratories C and D discontinued participation after the first year of the pilot study; therefore, the ability of the FP PT program to resolve this question is limited.

Negative bias was not confined to laboratories C and D; laboratory B also exhibited a substantial negative bias. In all three laboratories, the bias was more pronounced at higher blood lead concentrations. However, because laboratories B and D also displayed large positive y-intercepts, the bias is not apparent at lower blood lead concentrations. This is illustrated by the predicted results seen in Table 1. Regression slopes significantly below unity are consistent with observations made by Yee and Holtrop and Verebey et al., who reported slopes of 0.83 and 0.88, respectively, when comparing paired venous and FP samples in their studies.

There are two noteworthy limitations of the FP Pilot PT Program. One is the lack of FP PT specimens exceeding 1.93 μmol/L (40 μg/dL). Although the National Blood Lead PT Program routinely distributes specimens at these higher concentrations, the cohort of lead-exposed workers from which the FP specimens are drawn is routinely monitored by the Occupational Safety and Health Administration, which diminishes the probability of obtaining specimens exceeding 1.93 μmol/L (40 μg/dL). Alternative blood sources will be sought to overcome this limitation. In addition, questions relating to the potential contamination of the FP sheets during production and the potential for contamination of FP specimens during blood collection and handling were beyond the scope of this study. Other areas for additional study remain. The FP PT Program is investigating the suitability of a nonhuman blood source, the long-term stability (>60 days) of spotted FP sheets, the effect of medication use by donors, and the possible influence of hematocrit and lipemia on FP testing results. These data will be reported at a later date. Consistent with all PT, the results reported here were obtained on known evaluation specimens. Blind specimen submission would provide additional valuable data on the reliability of these methods.

Based on the results of this study, we have obtained preliminary Health Care Financing Administration certification for the FP PT Program as a CLIA provider. This will provide FP laboratories with service comparable to that available for laboratories testing liquid blood and will also require them to conform, at a minimum, to CLIA standards required for other blood lead testing laboratories. It is notable that only five laboratories currently are enrolled in the FP PT Program. Although only laboratories A, B, and E currently offer FP testing commercially, this number is likely to grow.

In conclusion, an effective FP PT program has been developed to capably assist and monitor FP laboratories. Based on overt proficiency testing results, properly utilized FP testing methods can accurately measure blood lead concentration.

This work was supported by a grant from the Division of Environmental Health Laboratory Services of the CDC through the Wisconsin State and Community-based Childhood Lead Poisoning Prevention Program Grant H64/CCH506537-06. In addition, we wish to acknowledge the invaluable support of Fred Langer and James Belanger of the United Paper Workers Local No. 7322, Milwaukee, WI.

References