The Performance of Laboratories Analysing Heavy Metals in the Workplace Analysis Scheme for Proficiency (WASP)

PETER R. STACEY*

The Health and Safety Laboratory, Harpur Hill, Buxton SK17 9JN, UK

Received 10 August 2005; in final form 14 December 2005; published online 24 January 2006

This paper discusses the proficiency of laboratories analysing identical test samples representative of metals in air in the Workplace Analysis Scheme for Proficiency (WASP) and the variability of results from laboratories used to assess personal exposure in the workplace. Over 11 years, the performance of laboratories has significantly improved for lead, cadmium and chromium. Laboratories show better agreement when using inductively coupled plasma-atomic emission spectroscopy (ICP-AES) than flame atomic absorption spectroscopy (FAAS) when analysing chromium, but not when analysing lead and cadmium. Data from the proficiency testing scheme show that for a customer to have confidence that a measurement value does not exceed the workplace exposure limit the analytical result would have to be <134 µg m⁻³ for lead and 23 µg m⁻³ for cadmium for an 8 h sample and 129 µg m⁻³ for lead, 21 µg m⁻³ for cadmium and 429 µg m⁻³ for chromium for a 4 h sample (for samples with relatively soluble matrices), before considering the further uncertainties due to sampling. The performance criterion for satisfactory performance (±8.3%) was found to be appropriate as long as laboratories participated consecutively in all rounds. Estimates of the overall uncertainty of results from ICP-AES and FAAS analysis were shown to meet the criteria in EN482. Samples from some types of workplace environments are more difficult than the WASP test material to dissolve into solution, so it is expected that estimates of uncertainty are larger for less soluble analytes and matrices. WASP has now initiated a programme to help laboratories assess their performance with more complex matrices.

Keywords: analysis; cadmium; chromium; lead; occupational hygiene; performance; proficiency-testing; uncertainty

INTRODUCTION

Lead, cadmium and chromium are probably the most frequently encountered analyses of metals in air in occupational hygiene. The health effects of these metals, when inhaled, are well known (Clayton and Clayton, 1994). Occupational exposure to lead in the UK is regulated by the Control of Lead at Work regulations (Stationery Office Ltd, 2002), which requires employers to assess the risk to health from exposure, and where exposure in air is >0.15 mg m⁻³ (150 µg m⁻³) for lead compounds or 0.1 mg m⁻³ (100 µg m⁻³) for tetra alkyl lead, the employer is required to provide respiratory protective equipment and monitor exposure. In the UK, occupational exposure to cadmium and chromium is regulated by the Control of Substances Hazardous to Health regulations, 2004 (Stationery Office Ltd, 2004). The limits for exposure to cadmium and chromium are 25 and 500 µg m⁻³, as stipulated in EH 40 (HSE, 2005).

The understanding of the reliability of analytical data is important to anyone trying to interpret and make decisions based on such data. Where programmes are available, proficiency testing (PT) is an integral part of assessing and proving the reliability of exposure results. Every three months, the WASP scheme supplies to laboratories four mixed cellulose ester filters each with a known amount of lead, cadmium and chromium and two blanks. This paper investigates the variation of results between laboratories at each analyte level over a recent 10 rounds of the scheme (2000–2002) and compares it with the performance of laboratories from the first 10 rounds of the scheme (1989–1991). It also

*Author to whom correspondence should be addressed. Tel: +44 (0)1298 218554; fax: +44 (0)1298 218571; e-mail: Peter.Stacey@hsl.gov.uk
may be other explanations for their improvement. It is possible that the poorer laboratories have discontinued and been replaced by laboratories who are keen to improve their proficiency. Also, laboratories are using newer techniques such as ICP-AES that may provide better accuracy and precision.

The regression line for the more recent between-laboratory standard deviations on the chart in Fig. 1 indicates that at the workplace exposure limit (WEL) of 150 µg m\(^{-3}\) for lead (144 µg on a filter for an 8 h sample at 2 l min\(^{-1}\)) analysis results from different laboratories of identical samples would produce a range of exposure values between 134 and 167 µg m\(^{-3}\) on ~95% of occasions (assuming the central limit theorem applies and the interlaboratory results are distributed normally). Figure 1 shows that for a customer to have confidence that a measurement value does not exceed the WEL the analytical result would have to be <134 µg m\(^{-3}\) before taking into consideration the additional variability from sampling. However, the chart for lead shows that when the scheme first started the variation of results at the WEL was poorer (±21.24 µg for 95% confidence). The corresponding range for cadmium at the WEL of 25 µg m\(^{-3}\) is ±3.1 µg (1.96σ). Because most cadmium samples are taken over 4 h, giving 12 µg on the filter, the range of expected reported exposure concentrations at the WEL is 21–29 µg m\(^{-3}\), which is slightly wider than for a sample of 8 h duration.

For chromium, the workplace WEL value for an 8 h sample is outside the analytical range studied (480 µg on the filter) so the example chosen is for a 4 h sample at the WEL (240 µg on the filter). The results suggest a corresponding range of 429–570 µg m\(^{-3}\) on 95% of occasions. However, when the scheme first started the variation of results was ±60 µg; which is almost double the current variation and the reported air concentration would have ranged from 375 to 625 µg m\(^{-3}\). Hygienists must, therefore, consider the expected variation of results and the potential for analytical bias.

### ANALYTICAL TECHNIQUES

The values for standard deviation in the charts in Fig. 1 include many different techniques such as FAAS, graphite furnace atomic absorption spectroscopy (GFAAS), inductively coupled plasma-mass spectroscopy (ICP-MS) and ICP-AES. Some of these techniques are expected to have better instrumental precision than others and should reduce the overall uncertainty of analysis. The most popular analytical techniques used in the scheme are ICP-AES and FAAS. Figure 2 compares the between laboratory standard deviations obtained by the most recent data at each loading for those laboratories that indicated they used either ICP-AES or FAAS.
The charts in Fig. 2 show that on average ICP-AES generally gives better between-laboratory standard deviation figures than FAAS. The difference is significant for chromium, which may be a consideration in measuring exposure near the WEL, but not for lead and cadmium.

**ANALYTICAL METHODS**

A comprehensive questionnaire was distributed to all laboratories in 1991. Although the number of returned questionnaires limited the reliability of the results there was some evidence to suggest that a combination of a low number of calibration standards and less vigorous dissolution procedures will lead to poorer performance. Of the seven laboratories using FAAS with category 3 (poor performance) values that completed all the questions in the questionnaire, four laboratories used 4 or fewer standards with their calibrations. Of these four laboratories, three used acid strengths <20% and two of these had heating times of <15 min.

---

**Fig. 1.** Performance of laboratories in the WASP programme for lead, cadmium and chromium.
There has been some concern by participants using the ICP-AES technique that the assigned value used in the WASP scheme for chromium (average of participants’ results excluding extreme values) is influenced by the different performances of the various techniques. The regression line in Fig. 3 shows the agreement between the average ICP-AES value and the average FAAS value for chromium.

This regression line is very sensitive to detecting differences between the two techniques because of the surprisingly good correlation coefficient (0.998) and the large number of values used, which results in a small standard error. The difference between the two techniques is statistically significant, but in terms of micrograms it is very small, being 0.37 μg (1%) at 100 μg and 9.6 μg (3.2%) at 300 μg. The equation for the regression line between ICP-AES and WASP assigned value (data not shown) is: ICP-AES average
The performance of laboratories analysing heavy metals in the WASP

value = 1.022 \times (\text{WASP ‘assigned’ value}) \ - \ 1.839.

The difference between the two using this equation is +0.67 \mu g (1\%) at 100 \mu g and +3.99 \mu g (1.3\%) at 300 \mu g. These differences are small when compared with the loadings used in the scheme, so the technique used is unlikely to affect performance assessment.

ARE THE PERFORMANCE LIMITS APPROPRIATE?

The performance statistic used to assess laboratories in the WASP scheme is discussed fully by Jackson and West (1992), Stacey and Wright (2001) and HSL (2001). The PT score, known in the scheme as the running performance index (RPI), is the average of the sum of the squared error of all the results submitted over four rounds. The square root of the score is related to the percentage relative standard deviation (%RSD) but will underestimate it by 0.97 as the denominator is ‘n’ and not ‘n – 1’. Currently, laboratories obtaining an RPI value >65 (8.3% RSD) are classified as unsatisfactory. This upper limit was initially determined from data supplied by an internal HSE quality assurance program, and is based on the 95% confidence limits of the expected distribution of results for a laboratory with average performance after the analysis of 16 samples. The expectation is that the average error for 95% of results are within ±8.3% and this is evaluated in this paper by calculating the average standardized error for each participating laboratory and by comparing the criteria with the 2.5 and 97.5% quantiles of these values in each round. In a single round, for a single sample, 95% of all standardized errors are between ±20% (0.8–1.18), which is the about the same as the present result rejection limits used in the scheme, and ~80% of results are between ±8%. This may suggest that these WASP performance criteria are too stringent. However, performance in WASP is not assessed on a single sample, as this is not representative of a laboratory’s ‘true’ performance and takes into consideration 16 results from five rounds. Assuming a laboratory has only random error, the uncertainty of all the measurements should decrease as it participates in more rounds. Table 2 examines the 2.5 and 97.5% quantiles for the average cumulated error obtained by laboratories over the last 5 rounds of WASP studied for this paper. If the performance criteria are valid then 95% of the cumulative average standardized ratios should be between 0.917 and 1.083 after five rounds.

Table 2 shows that for lead and cadmium 95% of laboratories’ cumulative average error is between ±8.3% (0.917–1.083) after five rounds. The ratios for chromium show a slight bias towards under reporting, as 4 of the 56 values obtained from the average error over all five rounds are <0.917 and weight the 2.5 percentile towards 0.903. Since the WASP programme assesses the performance of a laboratory on the best four of the last five rounds of performance values the scheme can be confident that, in the majority of cases, 95% of the average accumulated errors after five rounds are within the current limits of ±8.3%. However, the number of laboratories that participate in all five rounds is smaller than the number of laboratories that participated in the first round in Table 2. For example, the first round was evaluated using all results from 80 laboratories and only 56 of these laboratories participated consecutively over the five rounds.

Table 2. 2.5 and 97.5% quantiles using the simple average method for the cumulated average error over five rounds

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Quantile (%)</th>
<th>Round 1</th>
<th>Round 2</th>
<th>Round 3</th>
<th>Round 4</th>
<th>Round 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leada</td>
<td>2.5</td>
<td>0.884</td>
<td>0.876</td>
<td>0.881</td>
<td>0.913</td>
<td>0.915</td>
</tr>
<tr>
<td></td>
<td>97.5</td>
<td>1.08</td>
<td>1.224</td>
<td>1.118</td>
<td>1.111</td>
<td>1.074</td>
</tr>
<tr>
<td>Cadmiuma</td>
<td>2.5</td>
<td>0.913</td>
<td>0.893</td>
<td>0.895</td>
<td>0.919</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td>97.5</td>
<td>1.105</td>
<td>1.100</td>
<td>1.291</td>
<td>1.234</td>
<td>1.074</td>
</tr>
<tr>
<td>Chromium</td>
<td>2.5</td>
<td>0.788</td>
<td>0.847</td>
<td>0.811</td>
<td>0.841</td>
<td>0.903</td>
</tr>
<tr>
<td></td>
<td>97.5</td>
<td>1.137</td>
<td>1.001</td>
<td>1.066</td>
<td>1.072</td>
<td>1.062</td>
</tr>
</tbody>
</table>

aExcludes results from one laboratory with very large positive error.

Fig. 3. Difference between average of ICP-AES and FAAS results for chromium.
The figures in Table 2 suggest that it is essential for a laboratory to maintain its participation in five consecutive rounds in order to achieve a satisfactory rating in the WASP scheme. At the moment, laboratories that miss a single round in five have their performance assessed on the last four rounds of results and will have a greater risk of being classified a category 3 (poor performance) laboratory.

WITHIN LABORATORY PRECISION

Estimation of repeatability

The traditional approach used to determine the repeatability precision of a method is to calculate the standard deviation of results at different measured levels and to report this as a RSD. This usually represents the achievable repeatability precision of the method under optimized conditions within a single laboratory. The WASP scheme does not usually distribute replicate filters so to calculate an estimate of the repeatability precision, it has to be assumed that the relationship between the precision and loading is constant and then calculate the precision between each of the four samples distributed in each round. This repeatability precision estimate is determined by calculating the mean error sum of squares of the standardized ratios of the four results supplied by each laboratory. The figures in Table 3 show the values obtained for the two main analytical techniques used by laboratories in the scheme in terms of %RSD from the assigned values. It represents the average instrumental precision obtained by laboratories analysing WASP test filters on the same day with the same calibration.

Estimation of intermediate precision

Laboratories often use an intermediate precision estimate to determine the analytical precision component of their uncertainty calculations. Two replicate samples are usually analysed by the same laboratory under repeatability conditions on different days. While the WASP scheme does not usually distribute duplicate samples, but occasionally a similar loading is distributed in another round and these results can be paired. If enough pairs at different loadings are available, the relationship between intermediate precision and loading can be characterized across an analytical range. The precision of the sample preparation is often <1%, so the extra uncertainty in an additional sample will not greatly influence any calculated values. Figure 4 shows the estimated intermediate precision figures calculated from the six pairs of results. In order to compensate for slight differences in loading value between each pair of results, the precision estimates were calculated from the normalized values.

Unsurprisingly, the RSD values for the within laboratory intermediate precision determined from the equations of the trend lines in the charts in Fig. 4 are almost twice the estimated repeatability precision values (Table 3). An alternative method for determining the intra-laboratory variation is described in other work (Flinchbaugh et al., 2001) where the inter-laboratory variation is divided by \( \sqrt{2} \) to remove one source of error. Verification of the values shown on the charts with values derived from the inter-laboratory variation shows good agreement for cadmium and chromium but less agreement for lead where the relationship between intermediate within laboratory precision and the measurement value is less well defined. The charts indicate that the major factor contributing to the variability of results is changes between rounds, such as the changes with calibration factors. These values can act as intra-laboratory measurement quality objectives, as suggested by Flinchbaugh et al.

UNCERTAINTY

Since the scheme was launched in 1989, international performance standards such as EN482 have been established (BSI, 1994). EN482 states that the expanded uncertainty for both sampling and analysis should not be greater than \( \pm 30\% \) around the exposure limit (WEL) and \( \pm 50\% \) at one-tenth to one-half the WEL. The ISO guide for uncertainty measurement (ISO, 1995), applied within an individual laboratory, involves the identification of the individual sources of random and systematic error with the aid of cause and effect charts. The estimation of the overall uncertainty for a method for the analysis of metals and metalloids in air using ICP-AES and following the ISO approach is discussed in detail in an HSL report (HSL, 2002).

The widely used method of using ANOVA to calculate reproducibility standard deviation \( \Delta_R \) (ISO, 1995) is only valid if the analytical methods used are very similar and it is demonstrated that bias is not appreciable. PT schemes are not designed to determine reproducibility, however, the inter-laboratory standard deviation (excluding extreme values) often provides a good estimate of
reproducibility and the bias between the ‘true’ values and the ‘assigned’ scheme values for the WASP samples should be very small if many laboratories are involved. The uncertainty of the analyses of laboratories using the same analytical technique can be estimated from the equations given for the trend lines in the charts in Fig. 2. To assess if laboratories meet the requirements of EN482 we need to assign a value for the variability of the sampling processes related to the measurement and assume that the performance of the sampling equipment always meets the requirements in international standards, because the WASP PT scheme does not include the sampling process. Following procedures in

Fig. 4. Intermediate within laboratory precision values.
other work (HSL, 2002), sampler efficiency and flow rate variability are given values and included in the uncertainty estimation. International performance requirements (BSI, 1997; BSI, 2002) have percentage values for the maximum allowable variation for sampler efficiency and flow rate and these were adopted for each sampling process.

**Expanded uncertainty**

The inter-laboratory standard deviation, sampler efficiency and flow rate variability are combined as the square root of the sum of the squares and multiplied by a coverage factor (2) to provide the expanded uncertainty estimates for 8 h sampling on WASP type samples for FAAS and ICP-AES in Table 4.

If the pumps function within international performance standards on 99% of occasions then all these values are within the requirements of EH482 and all laboratories in WASP with satisfactory performance will comply with these criteria for samples with similar matrices. These estimates do not take into consideration the spatial variation of the aerosol across the breathing zone of the worker, which should be determined with additional measurements over a representative time period if the results are significant. As expected, the predicted expanded uncertainty values of the two techniques are smaller for ICP-AES analysis.

**Analysis with real samples**

Not all working environments produce metal fumes or dusts that are as soluble as the compounds used to prepare the test samples in this proficiency-testing program. Therefore, these samples may not provide the best assessment of a laboratory’s performance when analysing samples from some types of workplace environments, such as welding fume. It is important that the samples provided by quality assurance schemes represent the types of metal compounds and matrices that are found in real samples since this helps provide a better assessment of a laboratory’s performance. This is especially true when analysing chromium in some types of welding fume since vigorous digestion methods are needed to obtain fully recovery of the material into solution (Butler and Howe, 1999). It should be expected that measurements on some types of welding fume samples have a larger uncertainty because of the difficulties digesting these matrices. WASP is now introducing these ‘realistic’ sample types to assess the extent of the problem and to help laboratories improve their performance in this area.

**CONCLUSIONS**

- Over the last 11 years the performance of laboratories has improved. This improvement is significant for lead, cadmium and chromium.
- The performance of laboratories using ICP-AES is not significantly different from those laboratories using FAAS when analysing lead and cadmium.
- The performance of laboratories using ICP-AES is significantly different from those laboratories using FAAS when analysing chromium on WASP test filters.
- The difference between the assigned value for chromium and the average of results provided by ICP-AES instrument users is small and would not affect the assessment of the performance of laboratories.
- The criteria for satisfactory performance (±8.3%) are appropriate, but laboratories must participate in five sequential rounds to be certain of meeting this on 95% of occasions.
- The uncertainty of these measurements meets the criteria stipulated in EN482 if it is assumed that bias is taken into consideration.
- It is expected that the overall uncertainty following the ISO procedure will be larger for measurements of dust and fume in some types of workplace samples where more vigorous digestion procedures are needed to recover the material into solution.
- The results presented in this paper demonstrate the expected variation of results for laboratories that are keen to monitor their performance. Results from laboratories that do not take action to assess their bias cannot prove that the results they produce are reliable or achieve the same performance as those that participate in PT programmes.
- Data from the PT scheme show that for a customer to have confidence that a measurement value does not exceed the WEL the analytical result would have to be $<134 \text{ g m}^{-3}$ for lead and $23 \text{ g m}^{-3}$ for cadmium for an 8 h sample and $129 \text{ g m}^{-3}$ for lead, $21 \text{ g m}^{-3}$ for cadmium and $429 \text{ g m}^{-3}$ for chromium for a 4 h sample.

**Table 4. Estimate of percentage expanded uncertainty for FAAS and ICP-AES**

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Amount (µg)</th>
<th>FAAS Analysis</th>
<th>ICP-AES Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lead</td>
<td>WEL 144</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>0.5 WEL 72</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>0.2 WEL 28</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>Cadmium</td>
<td>WEL 48</td>
<td>20</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>0.3 WEL 8</td>
<td>19</td>
<td>17</td>
</tr>
<tr>
<td>Chromium</td>
<td>0.5WEL 240</td>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>0.25 WEL 120</td>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>0.2 WEL 96</td>
<td>19</td>
<td>13</td>
</tr>
</tbody>
</table>

424 P. R. Stacey
(for samples with relatively soluble matrices). It is strongly advisable that further action be taken to investigate the exposure if a higher value is obtained (but below the WEL), since it is within the expected distribution of results of a sample at the WEL. Analysis using the ICP-AES technique may improve these figures for chromium determination.

REFERENCES


