NISTIR 6468

Interlaboratory Studies on the Analysis of Hair for Drugs of Abuse

United States Department of Commerce
Technology Administration
National Institute of Standards and Technology

ABOUT THE LAW ENFORCEMENT AND CORRECTIONS STANDARDS AND TESTING PROGRAM

The Law Enforcement and Corrections Standards and Testing Program is sponsored by the Office of Science and Technology of the National Institute of Justice (NIJ), U.S. Department of Justice. The program responds to the mandate of the Justice System Improvement Act of 1979, which created NIJ and directed it to encourage research and development to improve the criminal justice system and to disseminate the results to Federal, State, and local agencies.

The Law Enforcement and Corrections Standards and Testing Program is an applied research effort that determines the technological needs of justice system agencies, sets minimum performance standards for specific devices, tests commercially available equipment against those standards, and disseminates the standards and the test results to criminal justice agencies nationally and internationally.

The program operates through:

The *Law Enforcement and Corrections Technology Advisory Council* (LECTAC) consisting of nationally recognized criminal justice practitioners from Federal, State, and local agencies, which assesses technological needs and sets priorities for research programs and items to be evaluated and tested.

The *Office of Law Enforcement Standards* (OLES) at the National Institute of Standards and Technology, which develops voluntary national performance standards for compliance testing to ensure that individual items of equipment are suitable for use by criminal justice agencies. The standards are based upon laboratory testing and evaluation of representative samples of each item of equipment to determine the key attributes, develop test methods, and establish minimum performance requirements for each essential attribute. In addition to the highly technical standards, OLES also produces technical reports and user guidelines that explain in nontechnical terms the capabilities of available equipment.

The *National Law Enforcement and Corrections Technology Center (NLECTC),* operated by a grantee, which supervises a national compliance testing program conducted by independent laboratories. The standards developed by OLES serve as performance benchmarks against which commercial equipment is measured. The facilities, personnel, and testing capabilities of the independent laboratories are evaluated by OLES prior to testing each item of equipment, and OLES helps the NLECTC staff review and analyze data. Test results are published in Equipment Performance Reports designed to help justice system procurement officials make informed purchasing decisions.

Publications are available at no charge through the National Law Enforcement and Corrections Technology Center. Some documents are also available online through the Internet/World Wide Web. To request a document or additional information, call 800-248-2742 or 301-519-5060, or write:

National Law Enforcement and Corrections Technology Center
P.O. Box 1160
Rockville, MD 20849-1160
E-Mail: *asknlectc@nlectc.org*
World Wide Web address: *http://www.nlectc.org*

The National Institute of Justice is a component of the Office of Justice Programs, which also includes the Bureau of Justice Assistance, Bureau of Justice Statistics, Office of Juvenile Justice and Delinquency Prevention, and the Office for Victims of Crime.

NISTIR 6468

Interlaboratory Studies on the Analysis of Hair for Drugs of Abuse

Michael J. Welch and Lorna T. Sniegoski
Chemical Science and Technology Laboratory
Analytical Chemistry Division
National Institute of Standards and Technology
Gaithersburg, MD 20899

March 2000

U.S. DEPARTMENT OF COMMERCE
William M. Daley, Secretary

TECHNOLOGY ADMINISTRATION
Dr. Cheryl Shavers, Under Secretary of Commerce for Technology

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
Raymond G. Kammer, Director

National Institute of Justice

Jeremy Travis
Director

The technical effort to develop this report was conducted
under Interagency Agreement 94-IJ-R-004,
Project No. 97-002/IA.

This report was prepared by the Office of Law Enforcement
Standards (OLES) of the National Institute of Standards
and Technology (NIST) under the direction of
Alim A. Fatah, Program Manager
for Chemical Systems and Materials, and
Kathleen M. Higgins, Director of OLES.
The work resulting from this report was sponsored by the
National Institute of Justice (NIJ), David G. Boyd, Director,
Office of Science and Technology.

FOREWORD

The Office of Law Enforcement Standards (OLES) of the National Institute of Standards and Technology (NIST) furnishes technical support to the National Institute of Justice (NIJ) program to strengthen law enforcement and criminal justice in the United States. OLES's function is to conduct research that will assist law enforcement and criminal justice agencies in the selection and procurement of quality equipment.

OLES is: (1) Subjecting existing equipment to laboratory testing and evaluation, and (2) conducting research leading to the development of several series of documents, including national standards, user guides, and technical reports.

This document covers research conducted by OLES under the sponsorship of the National Institute of Justice. Additional reports as well as other documents are being issued under the OLES program in the areas of protective clothing and equipment, communications systems, emergency equipment, investigative aids, security systems, vehicles, weapons, and analytical techniques and standard reference materials used by the forensic community.

Technical comments and suggestions concerning this report are invited from all interested parties. They may be addressed to the Office of Law Enforcement Standards, National Institute of Standards and Technology, 100 Bureau Drive, Stop 8102, Gaithersburg, MD 20899-8102.

Kathleen M. Higgins, Director
Office of Law Enforcement Standards

CONTENTS

TABLES

FIGURES

COMMONLY USED SYMBOLS AND ABBREVIATIONS

A	ampere	H	henry	nm	nanometer
ac	alternating current	h	hour	No.	number
AM	amplitude modulation	hf	high frequency	o.d.	outside diameter
cd	candela	Hz	hertz (c/s)	Ω	ohm
cm	centimeter	i.d.	inside diameter	p.	page
CP	chemically pure	in	inch	Pa	pascal
c/s	cycle per second	IR	infrared	pe	probable error
d	day	J	joule	pp.	pages
dB	decibel	L	lambert	ppm	parts per million
dc	direct current	L	liter	qt	quart
°C	degree Celsius	lb	pound	rad	radian
°F	degree Fahrenheit	lbf	pound-force	rf	radio frequency
dia	diameter	lbf·in	pound-force inch	rh	relative humidity
emf	electromotive force	lm	lumen	s	second
eq	equation	ln	logarithm (base e)	SD	standard deviation
F	farad	log	logarithm (base 10)	sec.	section
fc	footcandle	M	molar	SWR	standing wave ratio
fig.	figure	m	meter	uhf	ultrahigh frequency
FM	frequency modulation	min	minute	UV	ultraviolet
ft	foot	mm	millimeter	V	volt
ft/s	foot per second	mph	miles per hour	vhf	very high frequency
g	acceleration	m/s	meter per second	W	watt
g	gram	N	newton	λ	wavelength
gr	grain	N·m	newton meter	wt	weight

area=unit^2 (e.g., ft^2, in^2, etc.); volume=unit^3 (e.g., ft^3, m^3, etc.)

PREFIXES

d	deci (10^{-1})	da	deka (10)
c	centi (10^{-2})	h	hecto (10^2)
m	milli (10^{-3})	k	kilo (10^3)
μ	micro (10^{-6})	M	mega (10^6)
n	nano (10^{-9})	G	giga (10^9)
p	pico (10^{-12})	T	tera (10^{12})

COMMON CONVERSIONS
(See ASTM E380)

0.30480 m =1ft	4.448222 N = lbf
2.54 cm = 1 in	1.355818 J =1 ft·lbf
0.4535924 kg = 1 lb	0.1129848 N·m = lbf·in
0.06479891g = 1gr	14.59390 N/m =1 lbf/ft
0.9463529 L = 1 qt	6894.757 Pa = 1 lbf/in^2
3600000 J = 1 kW·hr	1.609344 km/h = mph

Temperature: $T_{°C} = (T_{°F}-32)\times 5/9$

Temperature: $T_{°F} = (T_{°C}\times 9/5)+32$

INTERLABORATORY STUDIES ON THE ANALYSIS OF HAIR FOR DRUGS OF ABUSE

Michael J. Welch and Lorna T. Sniegoski, Analytical Chemistry Division
Chemical Science and Technology Laboratory, National Institute of Standards and Technology

1. INTRODUCTION

In 1990, the National Institute of Standards and Technology (NIST) began research into the analysis of hair for drugs of abuse[1][1]. Included in that research has been an ongoing study of how well laboratories could detect and measure drugs of abuse in hair [2,3,4]. The study has included seven exercises in which NIST has sent hair samples to participants who agreed to analyze them. The purposes of these exercises were: (1) to evaluate the state-of-the-art for laboratories analyzing hair for drugs of abuse, (2) to provide laboratories with feedback on their performance, (3) to investigate the relationship between drug levels in hair and analytical accuracy, and (4) to compare the efficacy of various approaches used for hair analysis. The exercises were open to anyone who wished to participate and were strictly voluntary. No fees were charged for participation nor were the laboratories reimbursed for their costs associated with participation. The participants did not know which hair samples had target analytes in them, but they knew which compounds might be present.

A total of 26 laboratories participated in one or more exercises. Only a single laboratory participated in all seven of the exercises, while seven participated in only one. The participants included 15 laboratories from the U.S. and 11 from other countries, mostly in Western Europe. Law enforcement organizations, commercial drug testing laboratories, and university and government research laboratories made up most of the participants. The laboratories were promised confidentiality. The results from a given laboratory were known only to themselves and to the NIST staff involved with the study. After each exercise, each participating laboratory in that exercise was sent a report in which its results were compared with the overall mean and standard deviation of the group and with the NIST results. Erroneous results were noted. These fell into three categories: reports of finding target analytes when they were not present in the samples (false-positives), failing to find target analytes that were present (false-negatives), and reporting quantitative results that greatly disagreed with the study mean values (outliers).

For each exercise, six to eight samples of hair were sent to each participant. Each sample was one of three types: (1) drug-free hair, (2) drug-free hair into which NIST soaked target compounds, and (3) hair from known drug users provided by the participants. Although hair

[1]Numbers in brackets refer to suggested readings in section 4.

from drug users would seem to be most appropriate for these studies, there are several factors that limited the use of such hair. NIST does not perform drug tests on hair of individuals, and therefore, has no direct access to hair from drug users. The amount of hair required to enable each participant to test for all of the target compounds is very large relative to what is typically collected in drug testing situations. The distribution of drugs in the hair of a drug user may not be uniform. Finally, there may be other compounds present in such hair that would interfere with some of the analytical methods being used, thus adding an uncontrollable variable to the study. Therefore, NIST elected to use clean, drug-free hair into which selected target compounds were soaked for many of the samples in the exercises.

The first two exercises included cocaine (COC), benzoylecgonine (BZE), and morphine (MOR) as the target analytes, and codeine (COD) was added for the third exercise. In subsequent exercises additional analytes were added. Participants in the seventh exercise had to test for 10 analytes including 6-monoacetylmorphine (6-MAM), cocaethylene (CE), amphetamine (AMP), methamphetamine (MET), phencyclidine (PCP), and tetrahydrocannabinol (THC) in addition to the four original analytes. Levels of the target analytes in the recent exercises have generally been much lower than they were in the initial exercises, thus providing a greater challenge for the laboratories.

2. EXPERIMENTAL METHODS

2.1 Selection Process

2.1.1 Recruitment of Participants

Letters of invitation to participate were sent to laboratories known to be investigating hair analysis for drugs of abuse. Some of the participants spread the word to their colleagues, thus increasing the number of participants. All laboratories that responded positively for a particular exercise were sent samples.

2.1.2 Determination of Which Analytes to Include in the Exercise

The analytes included were based upon feedback from participants in previous exercises and on capabilities at NIST for accurate measurements for a particular analyte.

2.2 Hair Samples

2.2.1 Preparation of Hair Samples with the Target Analytes at Realistic Levels

NIST has no direct source of drug users' hair. Occasionally, reasonably large hair samples from individual drug users were donated to NIST for use in the exercises. These hair samples were washed and cut into small segments. For most exercises, the quantity of hair required for each

laboratory to perform all of their tests was relatively large compared to what is typically available from a drug user. Consequently, many of the samples included in the exercises were drug-free hair samples obtained from volunteers at NIST and treated with solutions of target compounds. After considerable trial and error, conditions were developed that permitted incorporation of realistic levels of the various analytes in these hair samples. Generally, the analytes were dissolved at appropriate concentrations in solutions of dimethylsulfoxide (DMSO) with or without water added. For most of the exercises, the hair was cut into short segments (0.3 cm to 0.7 cm typically) and considerable effort was expended to mix the hair in a given sample to improve homogeneity. However, for two exercises, the samples included hair that was cryogenically ground to a fine powder. The resulting material is very homogeneous and easy to work with, but is not as realistic or challenging as analyzing hair segments. Each exercise also included one or two unspiked samples from drug-free individuals.

2.2.2 Analysis of the Samples to Ensure That Appropriate Levels Were Achieved

NIST used gas chromatography/mass spectrometry (GC/MS) methods with isotope dilution to measure the concentrations of the target analytes in the hair samples. If a hair material that had been soaked in a solution containing one or more target analytes did not have appropriate levels, the hair was soaked further. If the concentration was too low, the hair was soaked longer or in a higher concentration solution to raise the levels. If the concentration was too high, the hair was soaked in a drug-free solution of DMSO to extract some of the analyte. Once appropriate concentrations were achieved, more samples from the batch were analyzed by GC/MS to better measure the mean concentrations of the analytes.

2.2.3 Preparation and Shipment of Samples to Participants

Fortunately, the hair materials are stable at room temperature for at least several weeks, so special precautions in shipping are not necessary. Each hair sample was placed in a small glass, screw-capped vial and carefully packed to prevent damage during shipment.

2.3 Data Assessment

2.3.1 Compilation and Interpretation of Results

With up to 14 participants per exercise, six to eight samples per lab, and up to 10 target compounds, a large quantity of data was generated. The results were entered into a large spreadsheet with a column for each participant. Each sample had a section of rows, with each row assigned to a target compound. With this approach it was possible to sum the positives and negatives reported for each analyte in each sample and to calculate means and standard deviations. The laboratories provided information about the methods they used, thus permitting tests for correlation between methods used and results. For two of the exercises, a control material, NIST Reference Material (RM) 8448 Drugs of Abuse in Human Hair Segments, was

sent along with the NIST assigned values, and the participants were asked to analyze the control along with the unknowns and report their results on the control. From these data, performance on the control material could be correlated with performance on the unknowns. Another aspect of the data to be investigated was the relationship between the level of analyte in the hair and analytical performance. Such a relationship is important when a laboratory establishes its cutoff levels.

2.3.2 Providing Feedback to Each Laboratory on its Performance Versus the Group

After the results were compiled and analyzed for each exercise, the participants were sent a report that showed how they performed versus the group as a whole. For each analyte present in a sample, information was provided on the mean, range, and standard deviation as well as the number of positive and negative responses and the number of outliers. For the analytes not present in a sample, the numbers of positive and negative responses were recorded. Each participant received an individualized report that included its results compared with the summary results, but did not have the individual results from the other participants.

3. RESULTS

3.1 Results from the Three Most Recent Exercises

The first four exercises have been described in detail previously [2,3,4]. A summary of the last three exercises, 5, 6, and 7 is shown in table 1. For exercises 5 and 6, samples included drug users' hair, soaked hair, and drug-free hair. The analytes for each exercise are listed along with the overall performance of the participants for each exercise. Three of the samples in exercise 5 were cryogenically powdered and three were in the form of segments. For exercises 6 and 7, all of the samples were in the form of segments. For exercise 7, only soaked hair and drug-free hair were included. For the positive challenges, the ability of the laboratories to detect the analytes correlates with the mean levels in the hair. Figure 1 shows how the percentage of correctly identified analytes increases with the mean level. For analyte levels below 0.5 ng/mg of hair, positive results were reported only 68 percent of the time, while for levels above 2 ng/mg, positive results were reported 100 percent of the time. In contrast, quantitative results do not show any significant correlation with concentration, as shown in figure 2, where the relationship between the interlaboratory coefficients of variation (CVs) and concentration is plotted. As with previous exercises, the quantitative results exhibit considerable scatter. These CVs are calculated after outliers are excluded from the calculations. For these studies, outliers were defined as any positive result that was more than a factor of three different from the mean. In these exercises, slightly fewer than half of the participants reported one or more results that were considered to be outliers, with most of these having one or two outliers in an exercise. In exercise 7, one laboratory accounted for five outliers out of a total of 12. Most of the outliers reported are low results. Laboratories with above average incidence of low results that were outliers were

advised to carefully examine their procedures, particularly their extractions, to see if their recoveries were low.

The incidence of false-positives was very low in the last three exercises. There were 12 instances of a particular drug being reported in a sample when it was not present. For exercises 5, 6, and 7, 18 laboratories participated in one or more exercises and three laboratories accounted for all of the false-positives. Two of these had false-positives in exercise 5 while the third had false-positives in all three exercises, including all of the false-positives in exercise 7. Thus, the most recent data suggest that most laboratories are quite careful, but there are a few that must improve performance if they are to use hair analysis in their work. The particular laboratory with the high error rate was advised to carefully check their results to determine if they were misinterpreting interferences as the target compounds. If the target compounds were actually present in the chromatograms, the laboratory must have had a serious problem with cross contamination, and was advised to take appropriate action.

Both drug users' hair and fortified (soaked) hair were included in exercise 5. A comparison of the results is shown in table 2. All of the missed positives for the soaked samples were for analytes whose mean concentrations were below 0.5 ng/mg as were half of the missed positives

Table 1. Overall results from interlaboratory Exercises 5, 6, & 7

	EXERCISE 5	6	7
Number of Participants	14	12	12
Number of Samples	6	6	6
Analytes	COC BZE CE MOR COD MAM	COC BZE CE MOR COD MAM AMP MET	COC BZE CE MOR COD MAM AMP MET THC PCP

Positive Challenges	5		6		7	
	Number	*%*	*Number*	*%*	*Number*	*%*
Positive Responses	197	87.6	143	84.3	100	92.6
Negative Responses	28	12.4	27	15.7	8	7.4
Outliers	11	5.6	10	7.0	11	11.0

Negative Challenges	5		6		7	
	Number	*%*	*Number*	*%*	*Number*	*%*
Positive Responses	4	1.6	2	0.7	6	1.4
Negative Responses	245	98.4	282	99.3	432	98.6

Table 2. Comparison of results by hair type and form for Exercise 5

Hair type	% Positives correctly identified	% Negatives correctly identified	% Outliers
Drug users'	90.6	97.7	8
Fortified	84.9	100	3
Drug-free		98.7	
Hair form			
Segments	98.7	99.4	2
Powder	81.5	96.7	9

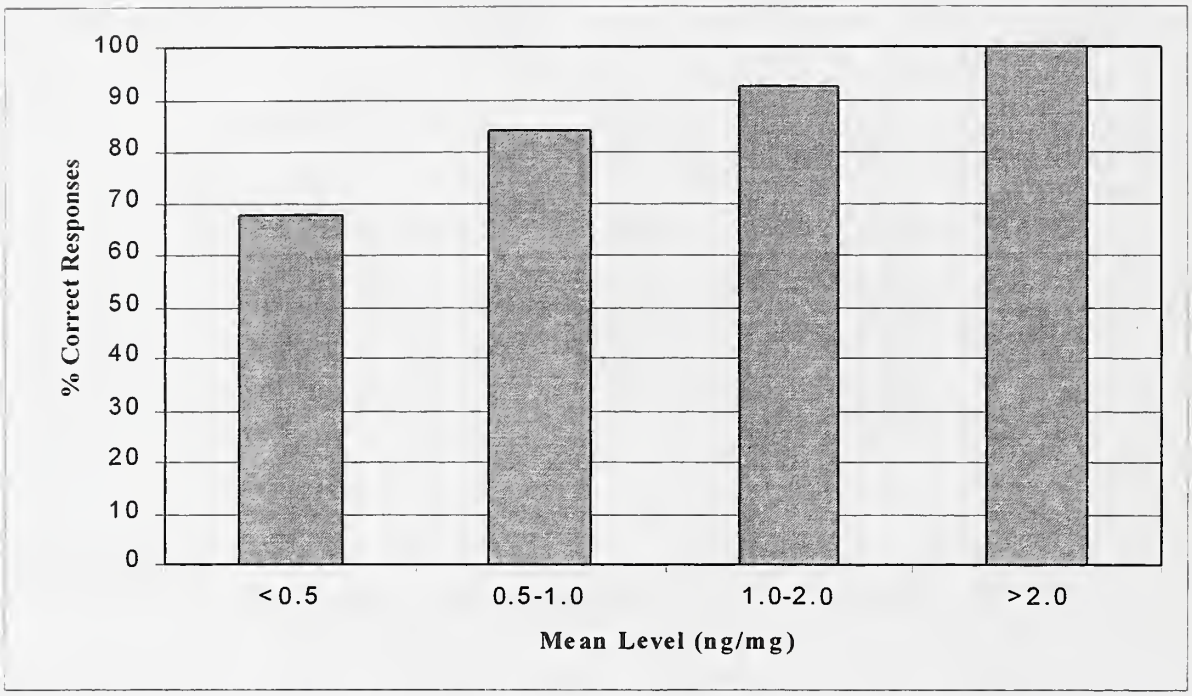

Figure 1. The relationship between mean analyte levels (in ng/mg) in hair samples and the ability of laboratories to correctly identify the presence of the analyte for exercises 5, 6, and 7.

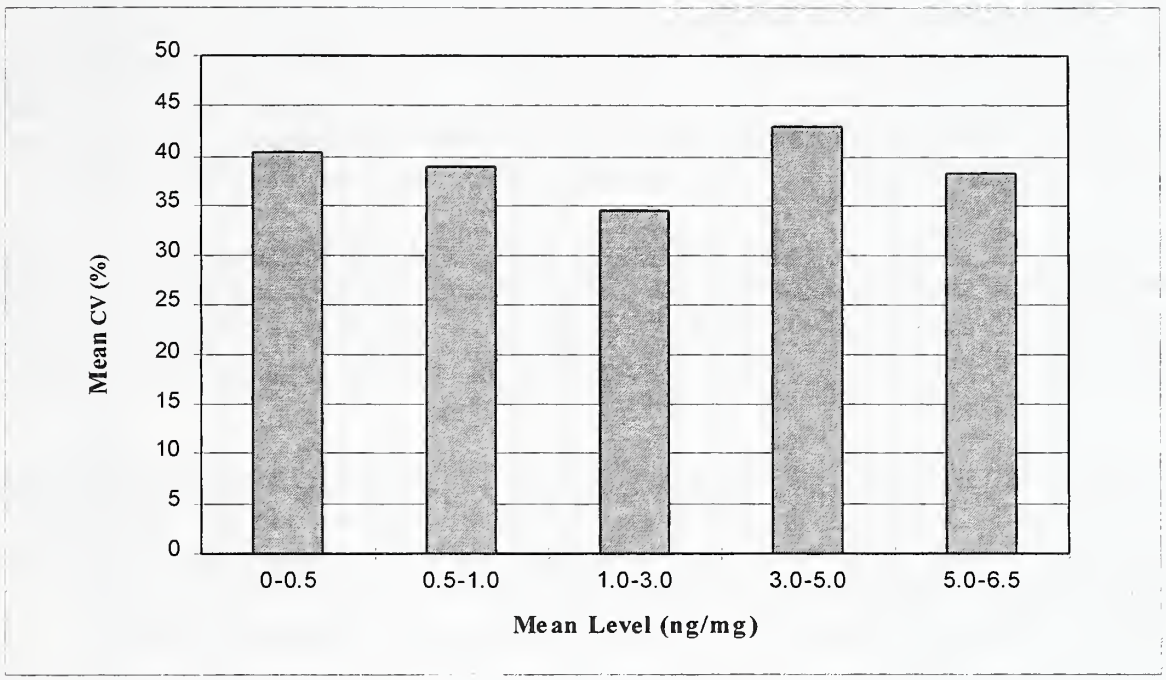

Figure 2. The relationship between mean analyte levels (in ng/mg) in hair samples and the interlaboratory precision for exercises 5,6, and 7.

on drug users' hair. This agrees with what was described above for the correlation between mean levels and the ability of the laboratories to detect the analytes. Otherwise, the performance was a little better for the soaked samples.

Exercise 5 also included both hair segments and powdered hair. These results are also summarized in table 2. Previous experience with these two forms suggested that results should be significantly better for the powdered hair. Analytes should be easier to extract from powdered hair and powdered hair samples should be more homogeneous than hair segment samples. However, for exercise 5, results were actually better for the hair segments. The average level of analytes in the hair segments was considerably higher than the average in the powdered hair. Thus one might expect fewer misses on positive challenges, but the results on the hair segments also had fewer false-positives and fewer outliers. Perhaps the fact that the laboratories have more experience with hair segments than with powdered hair accounts for the difference.

Exercise 6 was the first for which amphetamine (AMP) and methamphetamine (MET) were included and eight laboratories attempted to detect and measure each of them. Results were much better than had been seen for other analytes. There were no false-positives reported, no false-negatives, and only one outlier for the two samples that had these compounds present. The CVs averaged about 20 percent for each of these analytes.

Exercise 7 was the first exercise for which phencyclidine (PCP) and tetrahydrocannabinol (THC) were included and, again, eight laboratories attempted to detect and measure each of these analytes. Qualitatively, the results were quite good with only one missed positive between the two positive PCP samples and one missed positive for the one positive THC sample. Quantitatively, the interlaboratory CVs were quite large but there were no outliers.

As an example of the data reported back to the participants, the results for each hair sample in exercise 7 are summarized in table 3. The data in the "Your Results" column are not real, but is typical of what was observed. This column would have the results from the particular participant to whom the table was being sent. Outliers were counted as positives but were omitted from the calculation of the mean if they were more than a factor of three from the mean. The column headed by "mean w/o outliers" contains the mean results after these outliers were excluded. Thus each participant could compare its results with the overall mean and the NIST results to determine how good the agreement was. Although amphetamine and methamphetamine were included in the list of analytes to test for, they were not present in any of the samples. There were no false-positives among the nine laboratories that tested for amphetamine and methamphetamine. These results are omitted from the table to conserve space.

3.2 Composite Results From all Seven Exercises

To date, seven exercises have been completed. A summary of results on the seven exercises is shown in table 4. There has been no trend in the percentage of correct responses over the seven

Table 3. Example of results report sent to participants

Drugs in Hair RR 7				Overall Lab Results		Positive Specimens			Negative Specimens	
		Your	NIST	**Mean**	**Std Dev**	Positives	Negative	Number	Positives	Negative
Sample	Drug	**Result**	**Results**	w/o	w/o outliers	Reported	Reported	3X	Reported	Reported
1	COC	3.2	2.95	2.9	2.1	10	0	1		
	BZE	0.3	0.44	1.1	0.1	9	1	3		
	CE	**0**	0.32	0.3	0.3	7	1	2		
	MOR								0	11
	COD								0	10
	MAM								0	10
	THC								0	7
	PCP								0	7
2	COC								1	9
	BZE								1	9
	CE								0	8
	MOR								0	11
	COD								0	10
	MAM								0	10
	THC	X	2.81	1.6	1.1	6	1	1		
	PCP								0	7
3	COC								0	10
	BZE								1	9
	CE								0	8
	MOR								0	11
	COD								0	10
	MAM								0	10
	THC								0	7
	PCP	2.5	4.86	3.7	2.1	7	0	0		
4	COC								1	9
	BZE								1	9
	CE								0	8
	MOR								0	11
	COD								0	10
	MAM								0	10
	THC								0	7
	PCP								0	7
5	COC	6.2	2.09	1.8	0.4	10	0	2		
	BZE	0.4	0.22	0.5	0.2	10	0	1		
	CE	**0**	tr	0.1	0.1	6	2	3		
	MOR								0	11
	COD								0	10
	MAM								0	10
	THC								0	7
	PCP	0.8	1	0.7	0.5	6	1	1		
6	COC								0	10
	BZE	*0.5*							1	9
	CE								0	8
	MOR	4.4	6.9	3.6	1.3	11	0	1		
	COD	5.2	5.54	3.7	2.4	10	0	1		
	MAM	0.6	0.81	0.4	0.2	8	2	4		
	THC								0	7
	PCP								0	7

Pos. on Neg.	1
Neg. on Pos.	2
3x Outliers	2

Table 4. Overall results on positive and negative challenges for the seven exercises

Positive Challenges

	Responses		
Exercise	*Positive*	*Negative*	*% correct*
1	46	3	93.9
2	106	25	80.9
3	173	11	94.0
4	158	23	87.3
5	197	28	87.6
6	145	27	84.3
7	100	8	92.6
Overall	925	125	88.1

Negative Challenges

	Responses		
Exercise	*Positive*	*Negative*	*% correct*
1	0	63	100.0
2	5	72	93.5
3	5	179	97.3
4	22	323	93.6
5	4	245	98.4
6	2	282	99.3
7	6	432	98.6
Overall	44	1596	97.3

exercises. However, the challenges for the laboratories have generally become more difficult over the course of this study. The first exercise included only three analytes and these were present at very high concentrations. In subsequent exercises, more analytes were added and the analyte levels have decreased. In the most recent exercise, laboratories could test for up to 10 analytes and almost half of the mean concentrations were under 1 ng/mg.

The most important factors in good performance on these exercises are experience and adherence to good quality assurance procedures. This was clearly illustrated in exercise 4, when the laboratories analyzed NIST RM 8448 as a control [4]. The results showed a strong correlation between good performance on the reference material and good performance on the unknowns.

In general, the laboratories have performed at about the same level for all of the analytes. Table 5 shows the overall performance of the laboratories for the four compounds included in the largest number of exercises: cocaine, benzoylecgonine, morphine, and codeine. The results for exercises 2 through 7 were included in this table; exercise 1 was not, because cocaine and benzoylecgonine data were combined for this exercise to better compare results from different methods. The data in table 5 demonstrate that the overall performance is very similar for the different analytes.

As shown in table 5, the frequency of false-positives is also not related to the analytes. Figure 3 shows the combined results for COC, BZE, MOR, and COD for the number of false-positives as a function of the level that was reported. If the laboratories use a cutoff value, some of the false-positives seen in this study would probably not be reported as positives in actual drug testing situations. More than half of the false-positives reported levels below 1 ng/mg, so a cutoff at this level would reduce the number of false-positives by more than a factor of two. Most of the participants have not reported any false-positives. An average of 18 percent of the participants in a particular exercise have reported one or more false-positives, ranging from 0 percent for exercise 1 to 36 percent for exercise 4.

As shown in figure 4, there is no correlation between concentration and the rate of outliers. These results demonstrate that many of the laboratories have serious problems with their quantitative analysis not related to the limit of quantitation.

For both isolating the analytes from the hair matrix and for the analysis of the extracts, no single analytical approach has been found to be clearly superior. Participants in these studies have used three principal approaches, acid extraction, methanol extraction, and enzyme digestion for removing drugs from the hair matrix. For each of these approaches, some laboratories have consistently performed well, while others have been less successful, although error rates, particularly very low quantitative data, are significantly higher for laboratories using methanol extractions. For analysis of the extracts, GC/MS has been the approach used by most of the laboratories, with again a wide range in the quality of results received. A few laboratories have used other techniques for measurement. Good results were obtained by one laboratory

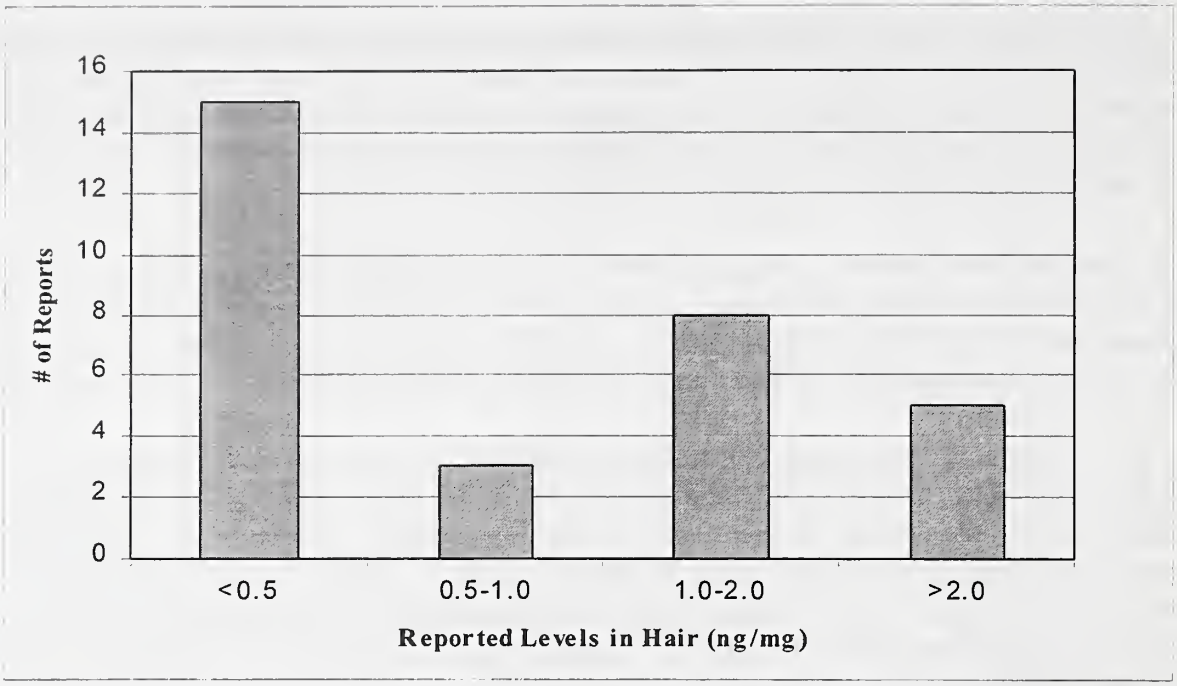

Figure 3. The relationship between reported analyte levels (in ng/mg) and the number of false-positives reported. Only data for cocaine, benzoylecgonine, morphine, and codeine in exercises 2 through 7 are included.

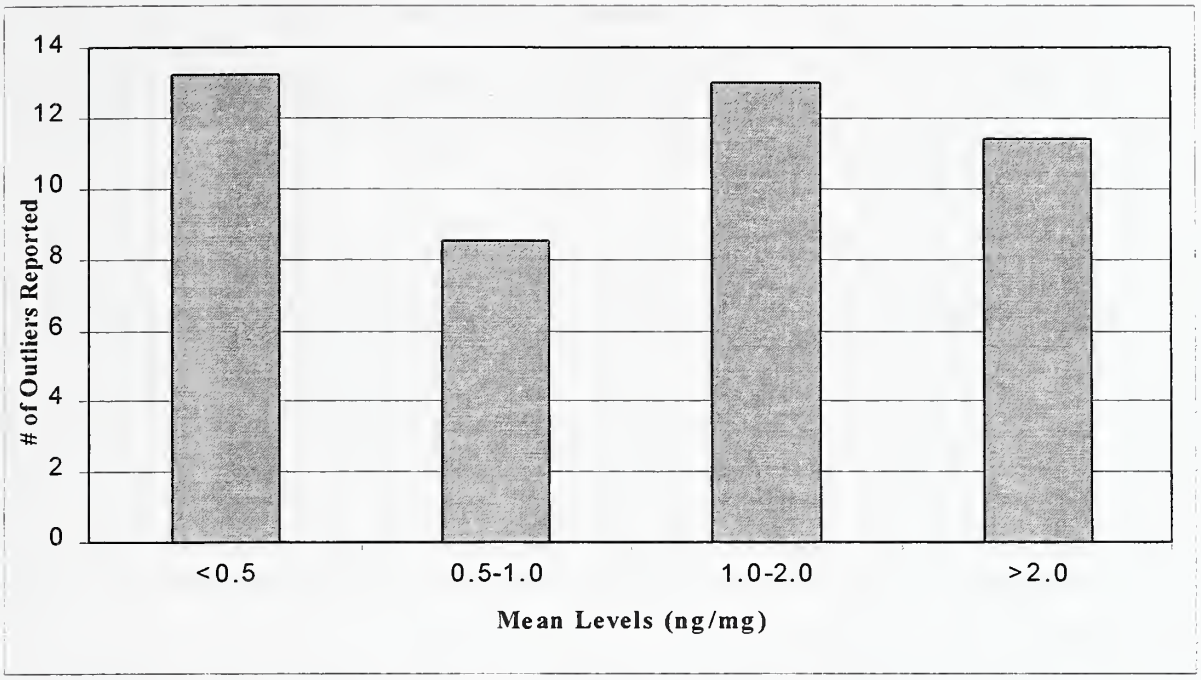

Figure 4. The relationship between mean analyte levels (in ng/mg) in hair samples and results that were outliers. Only data for cocaine, benzoylecgonine, morphine, and codeine in exercises 2 through 7 are included.

Table 5. Laboratory performance by analyte

	COC	BZE	MOR	COD
Correct Positives	88.3	86.3	89.4	88.1
Correct Negatives	96.4	96.6	96.1	98.8
% Outliers	14.0	11.2	9.3	9.3

performing tandem mass spectrometry (MS/MS) after extraction of analytes from the hair and further processing of the extracts. In contrast, the worst performance observed in any exercise involved a laboratory experimenting with direct MS/MS analysis of intact hair. Liquid chromatography with electrochemical detection has also been used with mixed results.

3.2 Future Work

Participation in a proficiency testing program is an important part of a laboratory's effort to evaluate its performance and to demonstrate its commitment to monitoring and improving the quality of its work. This study has served as a means for laboratories to assess the quality of their analyses. However, the focus was not on proficiency testing, but rather to assess the overall quality of laboratory measurements of drugs of abuse in hair. Laboratories and their customers would benefit from a formal proficiency testing program, involving at least two regularly scheduled exercises per year. Such a program should be a full-time activity for an organization with experience in proficiency testing programs for drug testing. NIST is not the organization to operate such a program, but could provide quality assurance for the program.

An important aspect of quality assurance that would fit with NIST strengths and interests would be to develop new Standard Reference Materials (SRMs) for drugs of abuse in hair that laboratories could use to evaluate the accuracy of their methods. Several years ago, NIST developed two reference materials (RM 8448 and 8449) that were human hair into which cocaine, benzoylecgonine, morphine, and codeine were soaked [5]. Concentrations were determined by GC/MS. As noted previously, good performance on control materials such as these RMs correlates well with performance on unknown specimens [4]. These materials are now out of stock and resources have not been available for a renewal of these materials. The next generation will be SRMs containing more analytes that are certified using two independent analytical methods. These reference materials will help laboratories assess and demonstrate the accuracy of their methods and help assure public confidence in forensic measurements.

4. SUGGESTED READINGS

[1] Welch, M.J., L.T. Sniegoski, C.C. Allgood, and M. Habram, Hair Analysis for Drugs of Abuse: Evaluation of Analytical Methods, Environmental Issues, and Development of Reference Materials, *J. Anal. Tox.,* 17 (1993) 389-396.

[2] Welch, M.J., L.T. Sniegoski, and C.C. Allgood, Interlaboratory Comparison Studies on the Analysis of Hair for Drugs of Abuse, *Forensic Sci. Int.,* 63 (1993) 295-303.

[3] Welch, M.J. and L.T. Sniegoski, "Interlaboratory Studies on the Analysis of Hair for Drugs of Abuse - Results from Three Exercises," *Hair Testing for Drugs of Abuse: International Research on Standards and Technology, NIH Publication No. 95-3727* (1995) 133-147.

[4] Sniegoski, L.T. and M.J. Welch, "Interlaboratory Studies on the Analysis of Hair for Drugs of Abuse: Results from the Fourth Exercise," *J. Anal. Tox.* 20 (1996) 242-247.

[5] Welch, M.J., P. Ellerbe, S.S.-C. Tai, R.G. Christensen, L.T. Sniegoski, L.C. Sander, and C.S. Phinney, "NIST Reference Materials to Support Accuracy in Drug Testing," *Fresenius J. Anal. Chem.*, 350 (1995) 4730-4734.