

HAIRVEQ 2006: Evolution of laboratories' performance after different educational actions

M. Ventura^{a,b}, C. Stramesi^c, S. Pichini^{d,*}, R. Ventura^{a,e}, M. Pujadas^a,
R. Di Giovannandrea^d, P. Zuccaro^d, R. Pacifici^d, K. Langohr^{a,f}, R. de la Torre^{a,e}

^a *Unitat de Recerca en Farmacologia, IMIM-Hospital del Mar, Parc de Recerca Biomèdica de Barcelona, Barcelona, Spain*

^b *UDIMAS, Universitat Autònoma de Barcelona, Barcelona, Spain*

^c *Dipartimento di Medicina Legale e Sanità Pubblica, Università di Pavia, Italy*

^d *Department of Therapeutic Research and Medicines Evaluation, Istituto Superiore di Sanità, V.le Regina Elena 299, 00161 Rome, Italy*

^e *CEXS, Universitat Pompeu Fabra, Barcelona, Spain*

^f *Departament d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya, Spain*

Received 14 May 2007; accepted 27 June 2007

Available online 5 November 2007

Abstract

HAIRVEQ is a proficiency testing program for hair analysis of illicit drugs organized by the Istituto Superiore di Sanità (Rome, Italy) and the Institut Municipal d'Investigació Mèdica (Barcelona, Spain). The aim of the three exercises performed in 2006 was the evaluation of 32 laboratories' performance when analyzing the same hair sample containing opiates, cocaine and methadone, after carrying out some specific educational interventions.

In the first round, the sample was sent to be analyzed following laboratory routine methodology. In the second round, standard operating procedures (SOP) for hair testing including sample preparation, method validation and qualitative and quantitative data evaluation, and an open hair sample for SOP training were also sent together with other hair samples including the one used for performance evaluation. After the second round, a workshop was held with participant laboratories to discuss methodological issues and interpretation of obtained results. An additional amount of open samples was distributed to the laboratories for implementing the SOPs. In the third round, the same unknown sample containing opiates, cocaine and methadone was resent for the final evaluation of laboratory performance. In the first round, 11 incorrect qualitative results (10 false negative and 1 false positive) were reported by seven laboratories (22%), in the second round, a reduction in the number of incorrect results was observed (4 false negatives and 1 false positive were reported by four laboratories, 13%) and in the third round, 5 false positives and 5 false negatives were reported by seven laboratories (22%). Concerning quantitative results, the scatter was similar between the three rounds and similar to the ones reported by other proficiency tests in hair analysis.

More educational actions should be addressed to a group of laboratories, which did not yet show satisfying qualitative and quantitative results.
© 2007 Elsevier Ireland Ltd. All rights reserved.

Keywords: Hair; Drugs of abuse; External proficiency testing

1. Introduction

The Istituto Superiore di Sanità of Rome, Italy, funded by the Italian Presidency of Ministers Council, has set up an external proficiency testing program (HAIRVEQ) in cooperation with the Institut Municipal d'Investigació Mèdica of Barcelona, Spain, to evaluate reliability in hair testing for drug abuse by

laboratories from the Italian National Health Service and Institutes of Forensic Medicine [1]. Since 2002 six different intercomparison exercises have been organized and 12 different hair samples have been distributed among the participant laboratories. The number of participants in the HAIRVEQ has increased from the initial 23 in 2002 to 32 laboratories in 2006. Results from the intercomparison exercises, organized from 2002 to 2005 have already been presented [1,2].

Particularly, the results obtained in 2005 showed that 4 years after the start of the program few laboratories still showed a satisfying performance for both qualitative and quantitative hair

* Corresponding author. Tel.: +39 06 49903682; fax: +39 06 49902016.

E-mail address: simona.pichini@iss.it (S. Pichini).

testing analysis. A unique source of error has not been identified as laboratories seem to apply correct digestion and extraction procedures and unsatisfying results cannot be attributed to the experience of the participating laboratories, nor to the analytical instrumentation or analytical method used [2]. Therefore, it was concluded that the most likely source of errors resided in method validation and evaluation of analytical results.

For this reason, at the beginning of 2006 standard operating procedures (SOP) on sample preparation, method validation and qualitative and quantitative data evaluation were prepared and distributed to the participating laboratories together with several milligrams of a hair sample with known concentrations of different analytes for training and implementing the SOPs. A workshop was held with participant laboratories at mid-2006 and additional amounts of open samples were distributed to the laboratories to keep on training and implementing the correct methodology.

In the meanwhile, the same hair sample was sent for three different intercomparison exercises with the aim of evaluating laboratories performance when analyzing the same hair sample,

before and after carrying out the above-reported specific educational interventions along with different exercises.

The present paper reports the results obtained in these three rounds and the evolution of the performance in hair testing for drugs of abuse by participating laboratories.

2. Experimental

2.1. Materials

Hair samples were obtained from drug addicts attending the detoxification unit of Hospital del Mar, Barcelona, Spain. The sample used for this specific evaluation was sent together with the other unknown ones. It was a real hair sample obtained from a drug consumer cut in short segments and homogenized containing opiates, cocaine and methadone. Four international reference laboratories verified sample content and homogeneity. Aliquots of 100 mg of hair were sent to each participant in each of the three different rounds.

In the first round, the sample was sent to be analyzed following laboratory routine methodology. In the second round, participating laboratories received, together with other hair samples including the one used for performance evaluation, an open sample with known analytes concentration and the recommended SOPs for training and adequately performing the analysis of opiates, cocaine, amphetamines and cannabinoids in hair. The SOPs contained information related

Table 1
Summary of standard operating procedures

	Opiates, cocaine and amphetamines	Cannabinoids
Sample preparation		
Washing	2 × 2 ml dichloromethane, vortex 1 min, centrifugation 5 min	2 × 2 ml dichloromethane, vortex 1 min, centrifugation 5 min
Amount	20–50 mg	20–50 mg
Control samples (calibration curve)		
Number	2 (C0, C2) or 6 (C0, C1, C2, C3, C4, C5)	2 (C0, C1) or 5 (C0, C1, C2, C3, C4)
Concentration (ng/mg of hair)	C0: 0 ng/mg, C1: 0.5 ng/mg, C2: 1 ng/mg, C3: 5 ng/mg, C4: 10 ng/mg, C5: 50 ng/mg	C0: 0 ng/mg, C1: 0.1 ng/mg, C2: 0.5 ng/mg, C3: 1 ng/mg, C4: 5 ng/mg
Sample analysis		
Internal standard	Deuterated or nalorphine for opiates and scopolamine for cocaine and amphetamines	Delta-8-THC or mephenamic acid
Digestion	1 ml 0.1M HCl, 45 °C, 18 h	1 ml 2 M NaOH, 45 °C, 30 min
Extraction ^a		
1 ml 0.1M sodium phosphate buffer pH = 6 + 50 µl of KOH 2M		–
Solid–liquid extraction (SPE)	Columns conditioning with methanol and phosphate buffer pH 6 Sample application Columns washing with deionized water, 0.1M HCl, methanol Elution with 2 ml dichloromethane:isopropanol:ammonium (80:20:2, v/v/v)	–
Liquid–liquid extraction (L–L)	Phosphate buffer pH 9.2 Extraction with 2 × 5 ml chloroform:isopropanol: <i>n</i> -heptane (50:17:33, v/v/v)	(For THC) Extraction with 5 ml hexane:ethylacetate (9:1, v/v), vortex 1 min, centrifugation, organic phase (For THC–COOH) Aqueous phase Addition of glacial acetic acid (pH 4) Extraction with 5 ml hexane:ethylacetate (9:1, v/v), vortex 1 min, centrifugation, organic phase
Organic phase evaporation		
Derivatization	50 µl MSTFA, 75 °C, 15 min	50 µl MSTFA, 75 °C, 15 min
Instrumental analysis	GC/MS	GC/MS

^a SPE or L–L extraction, one of them.

to sample preparation (washing, hair amount, internal standard), control samples (number, concentration and preparation), sample analysis (hair digestion, extraction, derivatization and instrumental analysis), qualitative and quantitative evaluation of results and recommended cut-off concentrations. See Table 1 for more information on sample preparation and analysis. Although very simplified, the proposed SOPs were the best cost-effective compromise to simultaneously analyze as much substances as possible. They were only an initial proposal of standardized procedures to be implemented once all the laboratories had been trained in the SOPs application. After the second round, a workshop was held with HAIRVEQ participants to discuss the results obtained in the second round after SOPs application, to solve any methodological and interpretative problem encountered. Additional amounts of open samples were distributed to the laboratories for implementing the SOPs during a 3-month period. In general, all the laboratories agreed in SOPs application, although they asked for a training time period. Finally, in the third round, the same sample was resent to the final evaluation of laboratory performance. Participating laboratories did not know the content of the sample used for performance evaluation in any of the three rounds.

2.2. Methods

Thirty-two laboratories participated in the three rounds of HAIRVEQ 2006, and four reference laboratories evaluated the qualitative and quantitative content of the specific hair sample (reported as robust mean \pm robust standard deviation, S.D.) [3] used for evaluation of laboratories performance. Hair sample (identified as S13, S15 and S19 in the first, second and third round, respectively) contained: 150.8 \pm 10.4 ng/mg hair cocaine (COC), 29.9 \pm 6.5 ng/mg hair benzoylecgonine (BZE), 4.9 \pm 1.0 ng/mg hair of 6-monoacetylmorphine (6-MAM), 3.6 \pm 0.2 ng/mg hair morphine (MOR), 0.9 \pm 0.1 ng/mg hair codeine (COD). Methadone (MET) was also present at a concentration of 82.4 ng/mg of hair, as identified by only one reference laboratory. These concentrations were taken as a reference to assess the content of the hair sample, which did not contain amphetamine (AP), methamphetamine (MA), nor any methylenedioxy derivative (3,4-methylenedioxy-methamphetamine MDMA, 3,4-methylenedioxyamphetamine MDA) or cannabinoids (Δ -9-tetrahydrocannabinol, THC and 11-nor- Δ -9-tetrahydrocannabinol-9-carboxylic acid THC-COOH).

For each analyte, dispersion and accuracy of the results reported by HAIRVEQ laboratories were measured. Dispersion was expressed as CV% (calculated using robust mean and robust standard deviation of participating laboratories, $CV\% = (\text{robust } sd_{\text{part}} / \text{robust mean}_{\text{part}}) \times 100$) and accuracy was measured as a standard error referred to the robust mean reported by reference laboratories ($ERR\% = (\text{robust mean}_{\text{part}} - \text{robust mean}_{\text{ref}}) / (\text{robust mean}_{\text{ref}}) \times 100$).

Evaluation of the quantitative performance of laboratories during 2006 was carried out calculating the Z-score values for the above-reported analytes in S13, S15 and S19. Two different Z-score values were calculated:

$Z\text{-score}^1 = (x - X) / \sigma$; where x was the individual laboratory's value for each target compound, X was the robust mean obtained by participating laboratories and σ the robust standard deviation obtained by participating laboratories [3].

$Z\text{-score}^2 = (x - X) / \sigma$; where x was the individual laboratory's value for each target compound; X was the robust mean value obtained by participating laboratories and σ was the expected relative standard deviation, measured according to Horwitz's equation [4,5].

3. Results and discussion

Qualitative results of samples in the 2006 HAIRVEQ rounds, including false negative (FN) and false positive (FP) results, are reported in Table 2. Both FN results and laboratories reporting FN decreased by a 50% (from 10 to 5 FN results and from 8 to 4 laboratories reporting a FN) after SOPs distribution and laboratory training with open samples. Differently, while in the first and in the second exercises only one FP result was reported, in the third exercise five FP results regarding AP, MA and MDA were given by three different laboratories. Although no reasonable explanation was found for the increase in the number of false positive results, it should be taken into account that three of them were reported for the same laboratory. Looking at the global qualitative results obtained by HAIRVEQ laboratories in the three rounds, it was observed that the 19 false negative results were reported by 9 of the 32 laboratories (28%) and the 7 false positives were reported by 5 of the 32 laboratories (16%). In 2002 round, incorrect results were reported by the 82% participating laboratories and in 2004 and 2005 rounds by 50% laboratories [1,2]. Hence, it can be said that in 2006 the majority of the participating laboratories had a satisfying qualitative performance, improving the overall methodology in terms of sensitivity and application of recommended cut-offs, although a group of laboratories persist in showing unsatisfying qualitative performance.

Regarding quantitative results, around 70% of HAIRVEQ laboratories reporting a result for these three rounds, quantified COC, BZE, MOR and COD, while 50% quantified 6-MAM and only 30% MET. Although amphetamine and cannabis

Table 2
Qualitative composition of samples, false negative and false positive results reported by HAIRVEQ laboratories

Sample ID	Qualitative content	6-MAM	MOR	COD	MET	COC	BZE	MDMA	MDA	AP	MA	THC	THC-COOH	Total	
False negative findings															
S13	6-MAM, MOR, COD, MET, COC, BZE	4 (F,G,J,L)	2 (F,J)	3 (E,F,I)	0	0	1 (L)	NA	NA	NA	NA	NA	NA	10 (E,F,G,I,J,L)	
S15	6-MAM, MOR, COD, MET, COC, BZE	1 (M)	0	0	1 (L)	1 (C)	1 (L)	NA	NA	NA	NA	NA	NA	4 (C,L,M)	
S19	6-MAM, MOR, COD, MET, COC, BZE	1 (M)	1 (D)	2 (D,E)	1 (J)	0	0	NA	NA	NA	NA	NA	NA	5 (D,E,J,M)	
False positive findings															
S13	6-MAM, MOR, COD, MET, COC, BZE	NA	NA	NA	NA	NA	NA	0	0	1 (B)	0	0	0	1 (B)	
S15	6-MAM, MOR, COD, MET, COC, BZE	NA	NA	NA	NA	NA	NA	0	1 (D)	0	0	0	0	1 (D)	
S19	6-MAM, MOR, COD, MET, COC, BZE	NA	NA	NA	NA	NA	NA	0	2 (A,H)	2 (A,K)	1 (A)	0	0	5 (A,H,K)	

NA: not applicable.

The code of laboratories reporting the erroneous qualitative results has been included in brackets.

Table 3

Dispersion and accuracy of the quantitative results reported by reference and HAIRVEQ laboratories for the three different 2006 rounds

	6-MAM		MOR		COD		MET		COC		BZE	
	REF ^a	HAIRVEQ ^b	REF ^a	HAIRVEQ ^b	REF ^a	HAIRVEQ ^b	REF ^a	HAIRVEQ ^b	REF ^a	HAIRVEQ ^b	REF ^a	HAIRVEQ ^b
S13												
CV% ^c	21.6	57.9	6.4	75.7	12.5	54.0	–	102.0	6.9	80.7	21.8	84.3
ERR% ^d	–	34.1	–	17.0	–	5.2	–	57.8	–	37.7	–	34.1
N ^e	4	16	4	27	4	15	1	13	4	28	4	23
S15												
CV% ^c	21.6	72.4	6.4	50.5	12.5	47.7	–	27.8	6.9	42.1	21.8	67.5
ERR% ^d	–	28.7	–	15.1	–	52.8	–	40.6	–	6.7	–	61.8
N ^e	4	17	4	22	4	17	1	8	4	22	4	22
S19												
CV% ^c	21.6	82.1	6.4	63.2	12.5	76.8	–	71.8	6.9	56.1	21.8	72.2
ERR% ^d	–	17.8	–	19.7	–	23.5	–	50.0	–	22.5	–	79.1
N ^e	4	19	4	24	4	20	1	9	4	23	4	23

^a Reference laboratories.^b HAIRVEQ participating laboratories.^c Coefficient of variation (%) = (robust sd_{part}/robust mean_{part}) × 100.^d Standard error (%) = ((robust mean_{part} – robust mean_{ref})/robust mean_{ref}) × 100.^e Number of laboratories reporting a quantitative result.

derivatives were absent in the sample object of this study, only 30% participating laboratories declared to quantify them.

Dispersion and accuracy of the results reported by reference and HAIRVEQ laboratories for the three different 2006 rounds were measured (Table 3). The dispersion in quantitative results (evaluated by the coefficient of variation, CV%) from HAIRVEQ laboratories remained quite high and in any case always higher than that obtained by reference laboratories, even if a slight improvement was noticed from the first to the third round for the majority of analytes. However, when comparing these dispersion values with those obtained in the previous years of the HAIRVEQ proficiency testing program, a significant two/three-fold decrease in CV% values can be observed [1,2] and the values obtained were similar to the ones reported in other proficiency tests on hair analysis for drugs of

abuse [4]. The accuracy (evaluated by standard error %) of the results improved only in case of 6-MAM and COC during the three different rounds, while it sensibly worsened in case of BZE.

The value of Z-score¹ and Z-score² for HAIRVEQ laboratories, which reported quantitative results in the three rounds of 2006, is shown in Figs. 1–4.

Z-score¹ (Figs. 1 and 2), calculated using robust mean and robust standard deviation of all values provided by participating laboratories is the recommended method of performance evaluation because no reported value is excluded and outlying data have less weight in the calculations. However, due to the high scatter generally found for concentration results in hair analysis and specifically in the HAIRVEQ program, the standard deviation calculated as for Z-score¹ is usually very

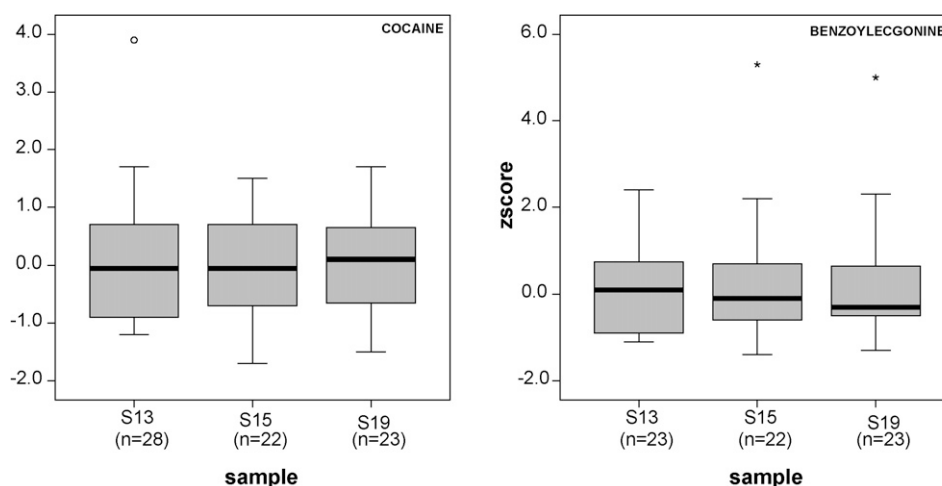


Fig. 1. Z-score¹ values of HAIRVEQ participating laboratories for cocaine and benzoylecgonine in the different rounds of 2006 (expressed as median value, central black line; values ranging from 25th to 75th percentile, grey box; the lowest and highest non-outlying values, the whiskers, and outliers, white circles and asterisks represented in box-plot diagrams).

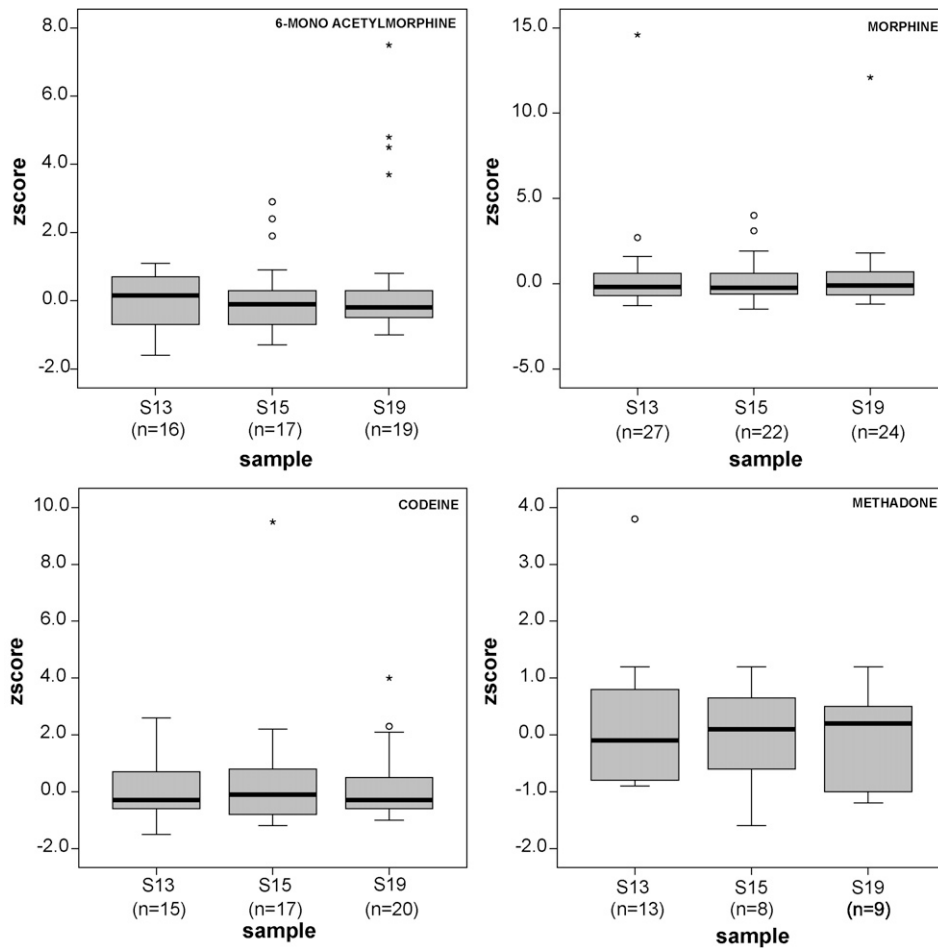


Fig. 2. Z-score¹ values of HAIRVEQ participating laboratories for opiates and methadone in the different rounds of 2006 (expressed as median value, central black line; values ranging from 25th to 75th percentile, grey box; the lowest and highest non-outlying values, the whiskers, and outliers, white circles and asterisks represented in box-plot diagrams).

high (in the three rounds of 2006 CVs% vary between 60 and 100%), with resulting satisfactory Z-score values for all participating laboratories (between -2 and 2). Taking into account that this was not logical, the Z-score² was also

calculated using the robust mean of all laboratories and the standard deviation measured according to Horwitz's equation, only dependent on the analyte concentration level (Figs. 3 and 4) [5,6]. This Z-score constitutes a value partially independent

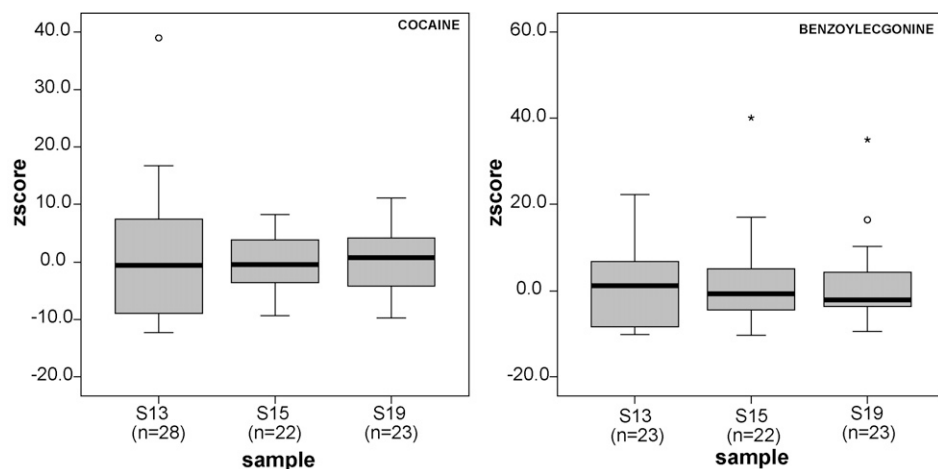


Fig. 3. Z-score² values for cocaine and benzoylecgonine calculated using robust mean of results reported by HAIRVEQ participating laboratories and sigma measured according to Horwitz equation. (Expressed as median value, central black line; values ranging from 25th to 75th percentile, grey box; the lowest and highest non-outlying values, the whiskers, and outliers, white circles and asterisks represented in box-plot diagrams).

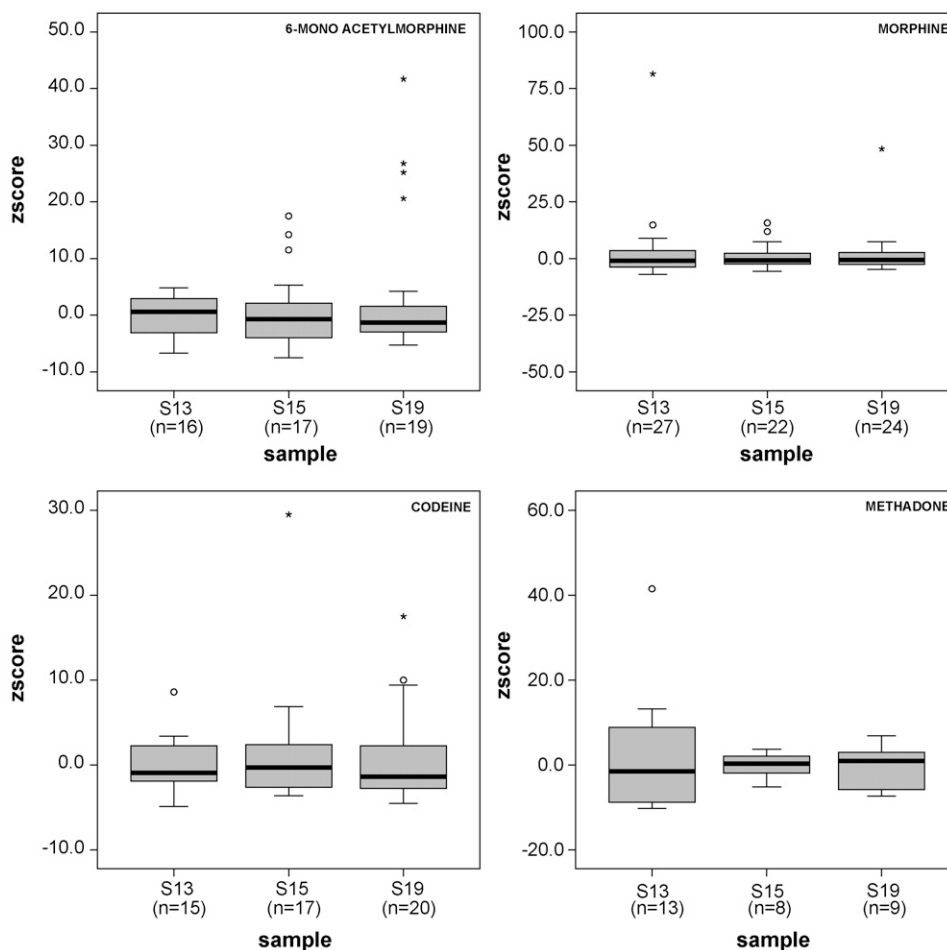


Fig. 4. Z-score² values for opiates and methadone calculated using robust mean of results reported by HAIRVEQ participating laboratories and sigma measured according to Horwitz equation. (Expressed as median value, central black line; values ranging from 25th to 75th percentile, grey box; the lowest and highest non-outlying values, the whiskers, and outliers, white circles and asterisks represented in box-plot diagrams).

from the type of performance obtained by participating laboratories for each target compound, since this parameter does not take into account standard deviation of laboratories but analyte concentration present in the sample under investigation. In hair analysis, spiked control samples cannot substitute real hair from drug users, because the drugs have to be extracted from the hair matrix. In any case, a more objective evaluation in the performance of hair testing for drugs of abuse by participating laboratories is obtained. Indeed, applying these calculations, only few laboratories showed satisfactory Z-score values. However, considering the Z-score² values of the last 2006 round, the median values in S19 laid between -1.3 and 1.0 (meaning satisfying results) for all analytes, apart for BZE (-2.1) and a moderate reduction in the values ranging from 25th to 75th percentile could be observed from S13 to S19 for all analytes but COD.

After the workshop, it was concluded that the unsatisfying qualitative and quantitative laboratories performance was mainly due to the incorrect sample preparation (e.g., not using internal standard), to the lack of complete validation of applied analytical methodologies (e.g., preparation of calibration curve) and to the incorrect qualitative and quantitative data evaluation (e.g., application of correct cut-offs or mass spectral

data interpretation). For this reason, it was strongly recommended to emphasize the following items:

- validation of the proposed SOPs and application to routine workload;
- use of a chromatographic technique (gas chromatography or liquid chromatography) coupled to mass spectrometry for samples analysis;
- use of internal standard and calibration curve to correctly identify and quantify the analytes eventually present in the samples;
- check of carry-over between samples and finally;
- use of three ions at least with the correct relative abundance ratios to identify different compound.

On one hand, it has to be acknowledged that, due to the educational actions carried out during 2006, qualitative performance significantly increased for the majority of HAIRVEQ laboratories with respect to the rounds held in the previous years. However, there is still a group of laboratories showing a non-satisfying qualitative performance. On the other, it has to be recognized a non-negligible scatter in quantitative results from

all the HAIRVEQ laboratories, similar to that observed in other proficiency programs in hair testing for illicit drugs [4,7–9].

In conclusion, we believe that the usefulness of an external proficiency testing program not only consists in assuring quality performance of participating laboratories, but also in providing educational tools for identifying and solving the encountered problems. Further actions of the HAIRVEQ program will focus on a specific training for mass spectrometric data interpretation and identifications of analytical pitfalls, specifically addressed to those laboratories not showing satisfying performance.

Acknowledgements

This investigation was supported by Dipartimento Nazionale per le politiche antidroga della Presidenza del Consiglio dei Ministri.

The organizers of HAIRVEQ are deeply grateful for the collaboration of the Society of Hair Testing and the following reference laboratories:

- Instituto Nacional de Toxicología y Ciencias Forenses, Sevilla, Spain (Dra. Carmen Jurado).
- Psychomedics Laboratories, USA (Dr. Michael Shaffer).
- Istituto di Medicina Legale Padova, Italy (Prof. Santo Davide Ferrara).

- Institut Municipal d'Investigació Mèdica, Pharmacology Unit, Barcelona, Spain (Dra. Rosa Ventura).

References

- [1] S. Pichini, M. Ventura, M. Pujadas, R. Ventura, M. Pellegrini, P. Zuccaro, R. Pacifici, R. de la Torre, Hairveq: an external quality control scheme for drugs of abuse analysis in hair, *Forensic Sci. Int.* 145 (2004) 109–115.
- [2] M. Ventura, S. Pichini, M. Pujadas, R. Ventura, R. Di Giovannandrea, P. Zuccaro, R. Pacifici, K. Langohr, C. Jurado, R. De la Torre, Four years' experience in external proficiency testing programs for hair testing of drugs of abuse in Italy (HAIRVEQ) and comparison with the society of hair testing program in 2005, *Ther. Drug Monit.* 29 (2007) 11–19.
- [3] International Standard ISO 13528: Statistical methods for use in proficiency testing by interlaboratory comparisons, (first ed., 2005-09-01).
- [4] C. Jurado, H. Sachs, Proficiency test for the analysis of hair for drugs of abuse, organized by the Society of Hair Testing, *Forensic Sci. Int.* 133 (2003) 175–178.
- [5] W. Horwitz, L.R. Kamps, K.W. Boyer, Qualità assurance in the analysis of foods for trace constituents, *J. Assoc. Off. Anal. Chem.* 63 (1980) 1344–1453.
- [6] W. Horwitz, R. Albert, The concept of uncertainty as applied to chemical measurements, *Analyst* 122 (1997) 615–617.
- [7] H. Sachs, Quality control by the Society of Hair Testing, *Forensic Sci. Int.* 84 (1997) 145–150.
- [8] M.J. Welch, L.T. Sniegoski, C.C. Allgood, Interlaboratory comparison studies on the analysis of hair for drugs of abuse, *Forensic Sci. Int.* 63 (1993) 295–303.
- [9] M. Deveaux, P. Kintz, J.P. Goullé, et al., The hair analysis proficiency testing program of the French Society of Analytical Toxicology, *Forensic Sci. Int.* 107 (2000) 389–394.