ORALVEQ: External quality assessment scheme of drugs of abuse in oral fluid
Results obtained in the first round performed in 2007

M. Ventura a,b, R. Ventura a,c, S. Pichini d,e, S. Leal a, P. Zuccaro d, R. Pacifici d, K. Langohr e,f, R. de la Torre c,e

a Grup de Recerca en Bioanàlisi i Serveis Analítics, Programa de Recerca en Neuropsicofarmacologia, IMIM-Hospital del Mar, Barcelona, Spain
b UDIMAS, Universitat Autònoma de Barcelona, Barcelona, Spain
c CEXS, Universitat Pompeu Fabra, Barcelona, Spain
d Department of Drug Research and Medicine Evaluation, Istituto Superiore di Sanità, V.le Regina Elena 299, 00161 Rome, Italy
e Grup de Recerca Clínica en Farmacologia Humana i Neurociències, Programa de Recerca en Neuropsicofarmacologia, IMIM-Hospital del Mar, Barcelona, Spain
f Departament d’Estadística i Investigació Operativa, Universitat Politécnica de Catalunya, Spain

1. Introduction

Oral fluid is a promising biological matrix for the analysis of drugs of abuse and, together with hair and sweat, a non-invasive specimen alternative, to blood and urine [1]. Oral fluid has several advantages over conventional biological fluids: its collection is simple and non-invasive and can be done under direct supervision to prevent sample adulteration, substitution and dilution. For this reason, in the last years, oral fluid has become an important analytical tool in clinical and forensic toxicology. A large number of studies have been conducted on the key issues in oral fluid drug testing [2–4], including the improvement and refinement of collection devices [5–8] and the development of more sensitive analytical techniques for onsite detection of drugs of abuse and subsequent confirmation by chromatographic techniques [9–10]. Furthermore, oral fluid testing for drugs of abuse has been proposed in Federal Workplace Drug Testing Programs Guidelines by the Substance Abuse and Mental Health Services Administration (SAMHSA) in the U.S. [11] and in Driving Under the Influence of Drugs (DUID) International Laws [12]. Subsequently, the number of international analytical laboratories routinely using oral fluid testing for drugs of abuse has increased in the last years.

Results provided by these laboratories need to be error-free since they can be used to take clinical and/or medical legal decisions. Hence, quality assurance of the results must be the prime objective of any analytical laboratory and the evaluation of laboratory performance through external quality assessment schemes (EQAS or proficiency testing) is recommended [13]. The
guidelines issued by SAMHSA, among initial mandatory requirements for a laboratory applicant to Workplace Drug Testing Programs, include the participation to EQAS [11].

To our knowledge, only one EQAS was organized in 2005 with samples collected with the Intercept oral fluid collection device [14].

During 2007, the Institut Municipal d’Investigació Mèdica IMIM–Hospital del Mar (Barcelona, Spain) in cooperation with the Istituto Superiore di Sanità (Rome, Italy) organized an external quality assessment scheme for drugs of abuse testing in oral fluid (ORALVEQ) to assess the reliability of analytical laboratories when analyzing drugs of abuse in oral fluid and the applied methodologies. Twenty-one laboratories (seventeen European and four American) participated in ORALVEQ, and the results reported are detailed and discussed in this paper.

2. Experiments

2.1. Materials

Three different samples (S1, S2 and S3) were prepared and sent to the participating laboratories. Samples were prepared with unstimulated oral fluid collected from healthy volunteers, pooled after centrifugation to remove the mucous part and analyzed to verify the absence of interfering substances using standardized methodologies [15,16]. S1 was a blank sample and S2 and S3 were prepared by addition of drugs at known concentrations to drug-free oral fluid. S2 contained 300 ng/ml of 6-monocetyl morphine (6-MAM), 800 ng/ml of morphine (MOR), 200 ng/ml of cocaine (COC) and 150 ng/ml of benzoylcodeine (BZE) and S3 contained 800 ng/ml of 3,4-methylendioxyamphetamine (MDMA) and 150 ng/ml of 3,4-methylendioxymethamphetamine (MDA). The analytes and their concentration present in the samples were selected according to oral fluid concentrations previously found in authentic samples from potentially intoxicated drivers [17]. The samples were prepared by spiking drug-free oral fluid (containing 0.1% sodium azide as preservative) with the analytes and diluted up to 50% with acidic buffer (0.1 M citric acid, 0.1 M sodium citrate, water, 33:17:50, v/v/v) at pH 4 [18]. The samples were mixed and distributed in 4 ml aliquots in Nunc-cryotubes™ (Nange Nunc International, Rochester, NY). For sample S2 and S3, two aliquots of eight tubes taken at the beginning, in the middle and at the end of the aliquoting procedure were analyzed using validated gas-chromatographic mass-spectrometric techniques [15,16] to verify samples’ content and homogeneity, and were found to be homogeneous. The rest of the tubes were stored at −20°C until sending them to the ORALVEQ laboratories at room temperature [18]. Three international reference laboratories verified samples content; they also found the different aliquots coming from tubes taken at the beginning, in the middle and at the end of the aliquoting procedure to be homogeneous for all analytes contained in S2 and S3.

2.2. Methods

Participating laboratories were informed about the groups of substances that could possibly be present in the samples (amphetamine (AMP)/methamphetamine (MA), MDMA/MDA, COC/BZE and/or MOR/6-MAM) and were suggested to apply the cut-off concentrations for oral fluid specimens recommended by SAMHSA [11].

Results were evaluated from a qualitative and quantitative point of view. The qualitative evaluation was done calculating false-negative and false-positive findings. The quantitative performance of each laboratory was carried out calculating the z-score value such as: 
\[
\text{z-score} = \frac{(X - \bar{X})}{\sigma},
\]
where \(X\) was the individual laboratory’s value, \(\bar{X}\) was the robust mean obtained by participating laboratories and \(\sigma\) the robust standard deviation obtained by participating laboratories. Robust mean and robust standard deviation were calculated following the International Standard ISO 13528: Statistical methods for use in proficiency testing by interlaboratory comparisons [19]. Criteria for classifying quantitative results on the basis of their z-score, were as follows: \(|z| < 2\) for satisfactory results, \(2 < |z| < 3\) for uncertain results and \(|z| \geq 3\) for unsatisfactory results. For each analyte, dispersion and accuracy of the results reported by participating laboratories were also measured. Dispersion was expressed as coefficient of variation (CV%) calculated using robust mean and robust standard deviation of participating laboratories and accuracy was measured as a standard error (ERR%) referred to the concentration reported by the reference laboratories.

3. Results

Twenty-one (seventeen European and four American) laboratories (including the three reference laboratories) participated in the first round of ORALVEQ. Regarding the analytical approach applied by laboratories, only 13 of the 21 laboratories (62%) performed screening of the samples using immunological methods. Enzyme-Linked Immunosorbent Assay—ELISA was the immunological method most applied by laboratories for all groups of substances (around 60%), followed by Cloned Enzyme Donor Immunoassay—CEDIA (around 30% for amphetamines and opiates and 23% for cocaine), Fluorescence Polarization Immunoassay—FPIA (around 8%) and Enzyme Multiplied Immunoassay—EMIT (around 8% for opiates and cocaine and not use for amphetamines). Approximately half of these laboratories used the screening cut-off concentrations recommended by SAMHSA [11], while the others applied either cut-off values suggested by assay manufacturer for urine samples or cut-off values established previously in their own laboratories during method development. All the 21 laboratories reported quantitative results for 6-MAM, MOR and COC, 19 laboratories for MDMA and BZE and 18 laboratories for MDA. One laboratory reported two values for each analyte using two different methodologies, rendering a total of 22 values for some analytes. A summary of the analytical approaches applied by the participating laboratories in the confirmation (identification/quantification) analysis is reported in Table 1.

Internal standards (deuterated analogues in 60% of the cases) were always used to quantify 6-MAM, MOR, COC and BZE. For amphetamines, all the participants except one used internal standard when analyzed the samples and 65% used the deuterated analogues of the analytes. Seventy percent of the labs used the SAMHSA cut-off concentrations to confirm amphetamines and around 50% for the rest of the analytes [11]. For both 6-MAM/MOR and COC/BZE approximately 70% of the labs used solid phase extraction (SPE), and 40% for amphetamines. 10% of the participating laboratories did not report the applied extraction procedure. Gas chromatography coupled to mass spectrometry (GC–MS) was the most used analytical technique (approximately 70% of the labs), followed by liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS) (approximately 30% of the labs) and liquid chromatography–mass spectrometry (LC–MS, 5% of the labs).

Some erroneous qualitative results were reported in the screening and confirmation analyses of the samples (Table 2). In the screening tests, performed by 13 laboratories, no false-negatives were found and a total of 10 false-positive results were reported by 2 laboratories: 5 false-positive results for S1 (the blank sample), 2 for S2 and 3 for S3. Regarding the confirmation tests, a
A total of 3 false-negative and 8 false-positive results were reported: 2 false-negatives in S2 and 1 in S3, 4 false-positives in S1 and 4 in S3 and of note, the 8 false-positive results were reported by 3 of the 21 laboratories (14%).

False-negative and false-positive results came from a total of 5 laboratories, with one laboratory giving either FN and FP results, other two laboratories false-negatives and the last two the false-positive results.

The majority of laboratories reported quantitative results (Table 3). ERR%, measured with respect to the concentration reported by reference laboratories, were approximately 5% for MOR and MDA and approximately 20% for 6-MAM, COC, BZE and MDMA. CVs% of participating laboratories, calculated using robust mean and robust standard deviation, were around 40% for all the analytes studied. In case of reference laboratories (n = 3) the CVs were below 10% for all analytes except for BZE (12.9%).

A high percentage (between 85 and 95%) of satisfactory results, in terms of z-score, was obtained for all analytes and no differences were observed between the global trend of the quantitative results obtained by laboratories using GC–MS and those using LC–MS/MS.

Concentrations reported by ORALVEQ laboratories (reference and participating laboratories) for 6-MAM and MOR, for COC and BZE and for MDMA and MDA are presented in Figs. 1–3, respectively.

Table 1
Summary of analytical approach applied by ORALVEQ laboratories in the confirmation (identification/quantification) analyses.

<table>
<thead>
<tr>
<th></th>
<th>Internal standard (% labs)</th>
<th>Extraction (% labs)</th>
<th>Analytical technique (% labs)</th>
<th>Cut-off concentration (% labs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>Deuterated</td>
<td>Non-deuterated</td>
<td>No</td>
</tr>
<tr>
<td>Amphetamine</td>
<td>5</td>
<td>65</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>Methamphetamine</td>
<td>5</td>
<td>65</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>MDMA</td>
<td>5</td>
<td>65</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>MDA</td>
<td>5</td>
<td>65</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>6-MAM</td>
<td>–</td>
<td>64</td>
<td>36</td>
<td>9</td>
</tr>
<tr>
<td>Morphine</td>
<td>–</td>
<td>64</td>
<td>36</td>
<td>9</td>
</tr>
<tr>
<td>Cocaine</td>
<td>–</td>
<td>59</td>
<td>41</td>
<td>9</td>
</tr>
<tr>
<td>Benzoyloxycodeine</td>
<td>–</td>
<td>59</td>
<td>41</td>
<td>9</td>
</tr>
</tbody>
</table>

Total of 3 false-negative and 8 false-positive results were reported: 2 false-negatives in S2 and 1 in S3, 4 false-positives in S1 and 4 in S3 and of note, the 8 false-positive results were reported by 3 of the 21 laboratories (14%).

False-negative and false-positive results came from a total of 5 laboratories, with one laboratory giving either FN and FP results, other two laboratories false-negatives and the last two the false-positive results.

The majority of laboratories reported quantitative results (Table 3). ERR%, measured with respect to the concentration reported by reference laboratories, were approximately 5% for MOR and MDA and approximately 20% for 6-MAM, COC, BZE and MDMA. CVs% of participating laboratories, calculated using robust mean and robust standard deviation, were around 40% for all the analytes studied. In case of reference laboratories (n = 3) the CVs were below 10% for all analytes except for BZE (12.9%).

A high percentage (between 85 and 95%) of satisfactory results, in terms of z-score, was obtained for all analytes and no differences were observed between the global trend of the quantitative results obtained by laboratories using GC–MS and those using LC–MS/MS.

Concentrations reported by ORALVEQ laboratories (reference and participating laboratories) for 6-MAM and MOR, for COC and BZE and for MDMA and MDA are presented in Figs. 1–3, respectively.

Table 2
Qualitative composition of samples and false-negative and false-positive findings.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Qualitative content</th>
<th>AMPHETAMINE CLASS</th>
<th>OPIATES</th>
<th>COCAINE</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SCR&lt;sup&gt;a&lt;/sup&gt;</td>
<td>AP</td>
<td>MA</td>
<td>MDMA</td>
<td>MDA</td>
</tr>
<tr>
<td>False-negative findings</td>
<td>S1</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>S2</td>
<td>6-MAM, MOR, COC, BZE</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>S3</td>
<td>MDMA, MDA</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>False-positive findings</td>
<td>S1</td>
<td>–</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S2</td>
<td>6-MAM, MOR, COC, BZE</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S3</td>
<td>MDMA, MDA</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<sup>a</sup> SCR: screening analysis.
<sup>b</sup> CONF: confirmation analysis: identification/quantification.

Table 3
Summary of quantitative results reported by ORALVEQ laboratories in S2 and S3.

<table>
<thead>
<tr>
<th>6-MAM</th>
<th>MOR</th>
<th>COC</th>
<th>BZE</th>
<th>MDMA</th>
<th>MDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Participant&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Reference&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Participant&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Reference&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Participant&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Mean&lt;sup&gt;c&lt;/sup&gt;</td>
<td>281.0</td>
<td>221.9</td>
<td>692.4</td>
<td>709.2</td>
<td>188.1</td>
</tr>
<tr>
<td>CVs&lt;sup&gt;d&lt;/sup&gt;</td>
<td>6.3</td>
<td>51.2</td>
<td>2.7</td>
<td>41.1</td>
<td>5.5</td>
</tr>
<tr>
<td>ERR&lt;sup&gt;e&lt;/sup&gt;</td>
<td>–</td>
<td>21.0</td>
<td>–</td>
<td>2.4</td>
<td>–</td>
</tr>
<tr>
<td>N&lt;sup&gt;f&lt;/sup&gt;</td>
<td>3</td>
<td>19</td>
<td>3</td>
<td>19</td>
<td>3</td>
</tr>
<tr>
<td>Satisfactory results&lt;sup&gt;g&lt;/sup&gt;</td>
<td>–</td>
<td>95</td>
<td>–</td>
<td>95</td>
<td>–</td>
</tr>
</tbody>
</table>

<sup>a</sup> Reference laboratories.
<sup>b</sup> Participating laboratories.
<sup>c</sup> Robust mean.
<sup>d</sup> Coefficient of variation (%) = (robust S.D./robust mean) × 100.
<sup>e</sup> Standard error (%) = ((robust mean – reference laboratories mean)/reference laboratories mean) × 100.
<sup>f</sup> Number of quantitative results.
<sup>g</sup> Percentage of satisfying results in terms of z-score (z ≤ 2).
4. Discussion

An important achievement of this first ORALVEQ round was that all the participating laboratories sent back the results. Of note, some of these laboratories do not routinely analyze oral fluid samples and had to re-apply to oral fluid and re-validate analytical methods developed and validated for screening and confirmation (identification/quantification) of drugs of abuse in other matrices, e.g. hair.

Only half of the laboratories performed screening tests prior to applying confirmatory analytical methods. This was probably caused by two factors: participating laboratories had been informed of the groups of substances that could be present in the samples and, at present time, the immunological methods available for the screening of drugs of abuse in oral fluid are scarce. This latter finding was also observed in the external quality assessment scheme (HAIRVEQ) we organized for drugs of abuse testing in hair [20].

All laboratories used cut-off concentrations to assess the presence or absence of a drug both in screening and in confirmation analyses; even if the cut-off concentrations recommended by SAMHSA were not applied for all analytes [11], the cut-off values established by local medical commissions or internally by laboratory itself were applied instead.

Fig. 1. Oral fluid concentrations reported by ORALVEQ laboratories for 6-MAM (A) and MOR (B) in S2 joined by analytical techniques (LC–MS, LC–MS/MS and GC–MS). Mean of reference laboratories indicated by a black line.

Fig. 2. Oral fluid concentrations reported by ORALVEQ laboratories for COC (A) and BZE (B) in S2 joined by analytical techniques (LC–MS, LC–MS/MS and GC–MS). Mean of reference laboratories indicated by a black line.
In the confirmation analyses, all laboratories used chromatographic techniques coupled to mass spectrometry. Inadequate methodologies, implying other than mass spectrometric detection reported for HAIRVEQ [21], were not used by ORALVEQ laboratories.

While the number of false-negatives reported was very low (only 3 false-negatives), it has to be recognized that a significant number of false-positive results was reported: 10 false-positives in screening tests and 8 in confirmation analysis. One possible explanation for the false-positive results reported in screening tests, as suggested by one of the reference Laboratories, could have been the interference of the acidic pH used to dilute oral fluid samples with the performance of one of the used immunoassay. The dilution with acidic buffer was necessary to avoid the hydrolysis of the ester bonds of cocaine and 6-MAM to benzyolarcjgonine and morphine, respectively [18,22]. The number of false-positive results in the confirmatory tests was quite high as well, but on a positive note, those erroneous results were reported by a reduced number of laboratories (3 out of 21). Both in case of false-negative and false-positive results, of note, these were reported by laboratories commonly engaged in drug testing in conventional and non-conventional matrices for clinical and medicolegal purposes. Moreover, before the starting of ORALVEQ first round, participating laboratories were specifically asked if they were prepared to analyze oral fluid samples.

A high number of laboratories reported quantitative results for all the analytes present in the samples. The accuracy of results was quite good (ERR% never above 20%), but at the same time the dispersion was quite high (CV% always around 40%), probably due to the presence of outliers; since applying robust statistics, no outliers were rejected but as an alternative to rejection they were given less weight. Indeed, CV% of reference laboratories were quite low for all analytes. Nevertheless, CV% obtained in this first round of ORALVEQ were similar to the ones found by Clarke and Wilson in the EQAS organized in 2005 with both oral fluid samples collected with the Intercept oral fluid collection device and samples prepared by dilution with the Intercept negative calibrator for oral fluid testing [14].

On the other hand, when comparing the accuracy and dispersion found for ORALVEQ samples with the ones obtained in proficiency testing programs for the analysis of drugs of abuse in another alternative matrix such as hair [21,23], the ERR% and the CV% found in ORALVEQ are much better. This is an expected, logical outcome taking into account that hair analyses require a more complex sample preparation (washing and digestion) than oral fluid.

Concerning z-score values, the high percentage of satisfactory results maybe due to the use of robust statistics, which takes into consideration all the results including outliers (even though with less weight), in the calculation of mean and S.D. [19].

5. Conclusion

This is the first external quality assessment scheme for drugs of abuse testing in oral fluid, which employs unstimulated oral fluid, treated only with sodium azide and acidic buffer.

Looking at the global results obtained in this proficiency testing for drugs of abuse in oral fluid, a satisfactory performance of laboratories has been obtained taking into account that some of the laboratories participating in ORALVEQ are not specifically experienced in analyzing oral fluid. Nevertheless, they should be prepared to apply also to oral fluid the analytical methods developed and validated for the screening and confirmation of drugs of abuse in other matrices.

In the screening tests, no false-negative results were reported and the false-positives could be related to the interference of the diluted oral fluid with one of the immunoassays used. With respect to confirmation analyses, the number of false-positive was considerable but they were reported by a reduced number of laboratories. Finally, concerning quantitative results, the scatter of results was quite high but this was due primarily to the few outlying values which were included since applying robust statistics.

Acknowledgements

The authors acknowledge the technical assistance of Meritxell Ventura Garcia and the collaboration of the following reference laboratories:
References


