Analytical Evaluation of Four On-Site Oral Fluid Drug Testing Devices

Sylvie Vanstechelman1*, Cristina Isalberti1, Trudy Van der Linden1, Kristof Pił2, Sara-Ann Legrand1 and Alain G. Verstraete3,4

1Department of Clinical Chemistry, Microbiology, and Immunology, Ghent University, De Pintelaan 185, 9000 Ghent, Belgium, 2Waters Corporation, Zellik, Belgium, 3Scientific Institute of Public Health, Juliette Wytsmanstraat 14, 1050 Brussels, Belgium, and 4Laboratory of Clinical Biology, Ghent University Hospital, De Pintelaan 185, 9000 Ghent, Belgium

* Author to whom correspondence should be addressed. Email: Sylvie.Vanstechelman@UGent.be.

The use of oral fluid (OF) as an alternative matrix for the detection of drugs of abuse has increased over the last decade, leading to the need for a rapid, simple, and reliable on-site OF testing device. Four on-site OF drug testing devices (Dra¨ger DrugTest 5000, Cozart DDS, Mavand Rapid STAT, and Innovacon OrAlert) were evaluated on 408 volunteers at drug treatment centers. UPLC–MS–MS results were used as reference to determine sensitivity, specificity and accuracy for each device, applying Belgian legal confirmation cutoffs for benzoylecgonine, cocaine, and THC (10 ng/mL); morphine and 6-acetylmorphine (5 ng/mL); and amphetamine and 3,4-methylenedioxyamphetamine (25 ng/mL). Sensitivity for cocaine was 50%, 50%, 27%, and 11% for DrugTest, OrAlert, Rapid STAT, and DDS 806, respectively. For opiates, sensitivities were 84%, 73%, 77%, and 65%, respectively. For THC, the sensitivities were 81%, 23%, 43%, and 28%, respectively. For amphetamines, the sensitivities were 75%, 33%, 17%, and 67%, respectively. Specificity was >88% for opiates and THC, >90% for amphetamines, and >97% for cocaine. All tests showed good specificity. DrugTest had the highest sensitivity, although it was still low for some analytes.

Introduction

Roadside surveys indicate that up to 15% of drivers screen positive for one or more drugs of abuse (1). The use of oral fluid (OF) as an alternative matrix for the detection of drugs of abuse has increased continuously over the last decade (2). Roadside OF testing to detect drugged driving has already been introduced in the legislation of some countries such as France and several Australian states (3, 4). In Belgium, from October 2010, drivers who appear impaired have to undergo a preliminary OF test. If the preliminary screening test is positive, a second OF sample or a blood sample is collected for confirmation analysis. If the confirmation analysis is also positive above the legal cutoffs, the driver will be prosecuted (5).

The aim of this study was to determine the reliability of four on-site OF drug testing devices: Dra¨ger DrugTest 5000, Cozart DDS, Mavand Rapid STAT, and Innovacon OrAlert. As a compromise between the very low DRUID cutoffs (6) and the different cutoffs of the devices, Belgian legal confirmation cutoffs were applied. These are 10 ng/mL for cocaine, benzoylecgonine, and THC, 5 ng/mL for opiates, and 25 ng/mL for amphetamines and 3,4-methylenedioxyamphetamine (MDMA) (5).

Materials and Methods

For the evaluation, the majority of the samples were collected in MSOC Ostend, a low threshold rehabilitation center for drug-dependent people. Smaller numbers of samples were also collected in a Psychiatric Emergency Intervention Centre and in a Drug Crisis Intervention Centre in Ghent (UPSIE). In all, 408 participants took part in the study. The study was approved by the ethics committee of Ghent University Hospital, and written informed consent was obtained from all volunteers (Belgian registration number B67020096426).

Samples were collected in the period July–November 2009, and each participant was asked to provide at least two OF samples. The first sample, collected with StatSure Saliva•sampler™, was used for confirmation analysis by ultra-performance liquid chromatography–tandem mass spectrometry (UPLC–MS–MS) after liquid–liquid extraction. Literature data have shown that some collection devices absorb most of the drugs, thus resulting in poor recoveries (7). This device, however, was chosen because of the good recovery of the analyzed drugs. [StatSure Saliva•sampler™ was the only device with recoveries greater than 80% for all the analytes in the study performed by Langel et al. (8).] A second sample was screened on-site with one of the devices. All devices are designed to detect cocaine, opiates, THC, and amphetamines. When possible, two different OF testing devices were tested on the same subject in the same session, leading to a total of 518 on-site tests performed. This depended on the amount of OF the participant could produce or if there was enough time left to perform more tests.

The DrugTest 5000 (Dra¨ger Safety, Lübeck, Germany) consists of a test cassette with an OF collector, a cartridge with buffer solution, and an analyzer. OF is collected by wiping the collector in the mouth and against the cheeks. When the indicator on the collector stick turns blue, it can be placed into the analyzer together with the buffer solution. Results are available within 8 min.

DDS 806 (Cozart Bioscience, Abingdon, U.K.) consists of a collector swab, a buffer bottle, a disposable test cartridge and a handheld instrument for result interpretation. The collection is similar to the collection of DrugTest 5000, but the mixing of the OF and buffer solution is not performed automatically. Results are available within 6 min.

RapidSTAT (Mavand Solutions, Mössingen, Germany) contains a collector stick with an aroma field, a buffer bottle, a test panel, and a reader. OrAlert (Innovacon, San Diego, CA) consists of a collector stick and a test device. After 3 min of rubbing the collector around in the mouth, the collector swab is squeezed into the test device. After 9 min, the results are available.

Results from the OF testing devices were interpreted visually in the case of OrAlert or using an automatic reader provided by the manufacturer for the other devices. The time between collection of the confirmation sample and the on-site test was...
generally less than 5 min. The OF samples for UPLC–MS–MS confirmation were stored at −20°C until analysis was performed with the methods described by Goessaert et al. (9). A sample was considered to be positive if any of the drugs mentioned in the Belgian legislation was present at a concentration above the cutoff mentioned in the law. For sensitivity and specificity, 95% confidence intervals (CI) were calculated by means of the modified Wald method (10). The formula that was used for this calculation is:

\[
p' = \frac{\text{# numerator} + 2}{\text{# denominator} + 4} = \frac{N + 2}{D + 4}
\]

\[
p' - 1.96 \sqrt{\frac{p' \times (1 - p')}{D + 4}} \text{ to } p' + 1.96 \sqrt{\frac{p' \times (1 - p')}{D + 4}}
\]

Results

Many problems were experienced while using the OrAlert device. The participants complained about the salty taste of the collection pad, and because there is no volume adequacy indicator, it was not clear if enough OF had been collected after 3 min. When too little OF is collected, the rest of the steps in the procedure can be jeopardized, and 15 tests failed after 3 min. When too little OF is collected, the rest of the collection pad, and because there is no volume adequacy indicator, it was not clear if enough OF had been collected. Ro¨hrich et al. (11) showed that too little fluid (buffer solution + OF) remains for confirmation analysis, which can compromise the confirmation. DDS test had fewer steps than RapidSTAT and it was easier to obtain the results. The main advantage of this test is the time of analysis only 6 min. DrugTest was the most practical because of simplicity of sample collection and analysis. Little training or experience is required to use this device. The effect of temperature may have an influence on the results. For example, according to the information provided, Mavand and Innovacon devices have to be used at room temperature, and, for DrugTest, the lowest operating temperature is 5°C, and many police roadside controls take place at night and in colder temperatures.

The evaluation of the qualitative results of the on-site tests was performed by the first author (SVS) in the majority of the cases. Thus, interindividual variability of the result interpretation was minimal. Moreover, all devices (except for OrAlert) were read with an electronic reader, which enhanced the objectivity and reproducibility of the screening results. Apart from OrAlert, with 13% failed tests, which were due to a lack of buffer solution and too little OF, the other tested devices met the Rosita-1 criteria (maximum 10% failures): 0% for DDS, 1% for DrugTest and 4% for RapidSTAT (12).

In Table I, sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) are given. The highest sensitivity for THC (81%, 62%–95%) (Figure 1), was observed with DrugTest, for which NPV and PPV were 64% and 98%, respectively. Overall, the other devices showed a sensitivity lower than 44%. Table II shows the number of non-zero concentrations of amphetamine, THC, cocaine, and morphine in OF sampled with the Statsure Saliva Sampler in subjects tested with the four onsite devices, as well as the minimum, maximum, median, and mean concentrations. A Kruskal-Wallis test on the drug concentrations in OF for the four different devices showed no significant differences (p > 0.05) in the drug concentrations, except for THC (p = 0.0117). For amphetamine, the difference was borderline (p = 0.0589).

For amphetamine detection, DrugTest also showed the highest sensitivity (75%), and DDS, OrAlert, and RapidSTAT had sensitivities of 67%, 33%, and 17%, respectively, with a wide CI (30–90%, 6–80%, 2–58%). The highest sensitivity for cocaine (50%) was achieved with DrugTest and OrAlert. RapidSTAT and

### Table I

<table>
<thead>
<tr>
<th>Target Substance</th>
<th>Device</th>
<th>Cutoffs (ng/mL)</th>
<th>n</th>
<th>Sens. (%)</th>
<th>Spec. (%)</th>
<th>Acc. (%)</th>
<th>Prev. (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cocaine/benzylecgonine</strong></td>
<td>DrugTest 20</td>
<td>137</td>
<td>50.0</td>
<td>99.2</td>
<td>94.9</td>
<td>8.8</td>
<td>93.2</td>
<td>90.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DDS 30</td>
<td>138</td>
<td>11.1</td>
<td>99.2</td>
<td>93.5</td>
<td>6.5</td>
<td>75.7</td>
<td>83.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RapidSTAT 12</td>
<td>133</td>
<td>27.3</td>
<td>97.5</td>
<td>91.7</td>
<td>8.3</td>
<td>70.7</td>
<td>86.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OrAlert 20</td>
<td>110</td>
<td>50.0</td>
<td>93.6</td>
<td>93.6</td>
<td>12.7</td>
<td>100</td>
<td>92.0</td>
<td></td>
</tr>
<tr>
<td><strong>Opiates</strong></td>
<td>DrugTest 20</td>
<td>137</td>
<td>84.1</td>
<td>91.8</td>
<td>92.0</td>
<td>64.2</td>
<td>22.2</td>
<td>99.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DDS 30</td>
<td>138</td>
<td>64.5</td>
<td>100</td>
<td>80.4</td>
<td>55.1</td>
<td>100</td>
<td>99.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RapidSTAT 25</td>
<td>133</td>
<td>75.0</td>
<td>96.1</td>
<td>84.2</td>
<td>60.9</td>
<td>35.6</td>
<td>99.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OrAlert 40</td>
<td>110</td>
<td>73.2</td>
<td>87.5</td>
<td>76.4</td>
<td>78.2</td>
<td>14.0</td>
<td>99.1</td>
<td></td>
</tr>
<tr>
<td><strong>THC</strong></td>
<td>DrugTest 5</td>
<td>137</td>
<td>80.8</td>
<td>95.5</td>
<td>90.0</td>
<td>24.8</td>
<td>98.0</td>
<td>64.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DDS 31</td>
<td>138</td>
<td>28.2</td>
<td>100</td>
<td>79.7</td>
<td>37.7</td>
<td>100</td>
<td>33.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RapidSTAT 15</td>
<td>133</td>
<td>43.3</td>
<td>88.3</td>
<td>78.2</td>
<td>22.5</td>
<td>91.2</td>
<td>36.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OrAlert 100</td>
<td>110</td>
<td>23.1</td>
<td>100</td>
<td>90.9</td>
<td>11.8</td>
<td>100</td>
<td>31.9</td>
<td></td>
</tr>
<tr>
<td><strong>Amphetamines</strong></td>
<td>DrugTest 50</td>
<td>137</td>
<td>75.0</td>
<td>96.5</td>
<td>92.0</td>
<td>5.8</td>
<td>100</td>
<td>94.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DDS 50</td>
<td>138</td>
<td>68.7</td>
<td>99.2</td>
<td>97.8</td>
<td>4.3</td>
<td>95.6</td>
<td>92.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RapidSTAT 25</td>
<td>133</td>
<td>16.7</td>
<td>96.9</td>
<td>93.2</td>
<td>4.5</td>
<td>56.6</td>
<td>82.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OrAlert 50</td>
<td>133</td>
<td>33.3</td>
<td>90.7</td>
<td>89.1</td>
<td>2.7</td>
<td>55.0</td>
<td>53.0</td>
<td></td>
</tr>
</tbody>
</table>

* The cutoffs mentioned are those claimed by the manufacturers for the different devices.

1. Sens. = Sensitivity = TP/(TP + FN); Spec. = Specificity = TN/(TN + FP); Acc. = Accuracy = (TP + TN)/(TP + TN + FP + FN); Prev. = Prevalence = (TP + FN)/(number of subjects); PPV = Positive Predictive Value = (sens*prev)/[(sens*prev) + (1 – spec*(1 – prev)); and NPV = Negative Predictive Value = spec*(1 – prev)/[(spec*(1 – prev) + prev*(1 – sens)]. TP = True Positive; TN = True Negative; FP = False Positive; and FN = False Negative.
Morphine, Cocaine, THC, Amphetamine

should be higher than 95% (9). Results can be influenced by a sensitivity and specificity should exceed 90% and accuracy criteria set in the Rosita-1 project, according to which both /C21 50% for all drug classes. Specificity was high for all drug classes with other drugs because of the low prevalence of amphetamines and THC, and 89% for amphetamine. Sensitivity and specificity are also influenced by the concentrations of drugs in the samples used for the evaluation. In a population with very high OF drug concentrations, a good sensitivity is easier to achieve. For opiates, the high sensitivity may be explained by a combination of factors. As can be seen in Figure 2, the concentrations of opiates were generally very high. Moreover, the concomitant presence of other opiates, such as codeine and dihydrocodeine, not included in the Belgian law or not analyzed in the confirmation method, may explain the good sensitivities. However, the good specificity illustrates that these drugs did not cause many false positives in this population.

In general, confirmation analysis revealed concentrations of cocaine that were mostly below the cutoffs set by the manufacturers of the devices (30 ng/mL cocaine for Dräger, 12, 20, and 30 ng/mL benzoylecgonine for Rapid Stat, Oralert, and Cozart, respectively) (Figure 2). This leads to an increased risk of having false-negative results and partially explains the low sensitivity for cocaine shown by all devices. The prevalence of opiates and cannabis was high, explaining the smaller CI.

The choice of study population has a large influence on the found results. The population in which we evaluated the devices, a population with a lot of positive subjects, is justified because it is very difficult to obtain a high number of positive results in normal traffic. Additionally, a large population of positive results makes our findings more accurate and precise, especially for cannabis and opiates, where the prevalence of positive subjects was high. To ascertain the usefulness of these on-site devices, PPV and NPV, based on the prevalence of the different types of drugs in the target population (13), were calculated. Keeping the study population of present research in mind, it was expected to find higher prevalence of drugs, with higher PPV and lower NPV compared to normal traffic, where the prevalence is lower. In contrast, the prevalence of amphetamines and cannabis in the studied population can be assumed to be lower than in a pre-selected roadside setting. This could compromise a comparison of the sensitivity for these drug classes with other studies, but the calculated confidence intervals do take this into account.

The fact that one sample was not used to test all four devices can be seen as a limitation. The distribution of concentrations (and interferences) in the samples used for each of the procedures could have been different, and thus biased the comparisons of sensitivity. For practical reasons (it is not easy to obtain enough OF from a subject who has recently taken drugs), this was not possible. The drug concentrations in the samples tested with the different devices are shown in Table II. The THC concentrations are higher with Cozart, but this device had a low sensitivity (28% for THC), so the higher concentrations in the oral fluid samples would rather introduce a bias towards a better sensitivity for this device.

Sensitivities found in the present study are lower than those found in the study by Wille et al. (14), where DrugTest reached a sensitivity of 100% for amphetamines and 93% for THC, and RapidSTAT reached 93% and 71%, respectively. However, they compared with confirmation results in blood, and other studies

**Discussion**

The results show that only DrugTest reached a sensitivity ≥50% for all drug classes. Specificity was high for all drug classes for each test. None of the devices, however, met the criteria set in the Rosita-1 project, according to which both sensitivity and specificity should exceed 90% and accuracy should be higher than 95% (9). Results can be influenced by a difference in recovery of the drug analytes with the four collection devices, but this should be investigated further. According to Langel et al. (8), the recovery results for StatSure Saliva® sampler™ are 86% for cocaine, 88% for morphine, 81% for codeine, 85% for THC, and 89% for amphetamine. Sensitivity and specificity are also influenced by the concentrations of drugs in the samples used for the evaluation. In a population with very high OF drug concentrations, a good sensitivity is easier to achieve. For opiates, the high sensitivity may be explained by a combination of factors. As can be seen in Figure 2, the concentrations of opiates were generally very high. Moreover, the concomitant presence of other opiates, such as codeine and dihydrocodeine, not included in the Belgian law or not analyzed in the confirmation method, may explain the good sensitivities. However, the good specificity illustrates that these drugs did not cause many false positives in this population.

In general, confirmation analysis revealed concentrations of cocaine that were mostly below the cutoffs set by the manufacturers of the devices (30 ng/mL cocaine for Dräger, 12, 20, and 30 ng/mL benzoylecgonine for Rapid Stat, Oralert, and Cozart, respectively) (Figure 2). This leads to an increased risk of having false-negative results and partially explains the low sensitivity for cocaine shown by all devices. The prevalence of opiates and cannabis was high, explaining the smaller CI.

The choice of study population has a large influence on the found results. The population in which we evaluated the devices, a population with a lot of positive subjects, is justified because it is very difficult to obtain a high number of positive results in normal traffic. Additionally, a large population of positive results makes our findings more accurate and precise, especially for cannabis and opiates, where the prevalence of positive subjects was high. To ascertain the usefulness of these on-site devices, PPV and NPV, based on the prevalence of the different types of drugs in the target population (13), were calculated. Keeping the study population of present research in mind, it was expected to find higher prevalence of drugs, with higher PPV and lower NPV compared to normal traffic, where the prevalence is lower. In contrast, the prevalence of amphetamines and cannabis in the studied population can be assumed to be lower than in a pre-selected roadside setting. This could compromise a comparison of the sensitivity for these drug classes with other studies, but the calculated confidence intervals do take this into account.

The fact that one sample was not used to test all four devices can be seen as a limitation. The distribution of concentrations (and interferences) in the samples used for each of the procedures could have been different, and thus biased the comparisons of sensitivity. For practical reasons (it is not easy to obtain enough OF from a subject who has recently taken drugs), this was not possible. The drug concentrations in the samples tested with the different devices are shown in Table II. The THC concentrations are higher with Cozart, but this device had a low sensitivity (28% for THC), so the higher concentrations in the oral fluid samples would rather introduce a bias towards a better sensitivity for this device.

Sensitivities found in the present study are lower than those found in the study by Wille et al. (14), where DrugTest reached a sensitivity of 100% for amphetamines and 93% for THC, and RapidSTAT reached 93% and 71%, respectively. However, they compared with confirmation results in blood, and other studies
have shown that there is a better agreement between OF on-site results and blood than there is with oral fluid (15). Röhrich et al. (11) found sensitivities of 94% and 85% for amphetamines and THC, respectively, when tested with RapidSTAT, while in our study these were only 17% and 43%. This may be due to the lower cutoffs set in this study. Also, unlike in the study performed by Röhrich et al. (11), the Mavand reader was always used to objectively interpret the results. Crouch et al. (16) also tested 10 different devices, and although one test, Drugwipe, showed promising results, it did not meet the study criteria for acceptable device performance (90% or greater sensitivity and specificity). All these previous studies used the manufacturer cutoffs, which are equal to or higher than the cutoffs in the Belgian Law, except for THC with the DrugTest (5 ng/mL). Walsh et al. (17) also evaluated 10 different devices with spiked OF and checked if the cutoff concentrations postulated by the manufacturer were achieved. For most tests, false negatives and false positives were detected for each drug type. Thus, it can be concluded that the devices were not able to live up to their expectations in terms of reliability. In most other studies, the sensitivity and specificity were too low; for example, the Drugwipe 5 gave unsatisfactory results for cannabis (resp. 52.5% and 91.2%) (18).

In the Rosita-2 study (performed between 2003 and 2005), the analytical evaluation of the cannabis tests showed sensitivity and specificity in the range of 0–74% and 70–100%, respectively. We definitely see an improvement in this study with 23–81% and 88–100%, respectively (15).

Conclusions
Among the four devices, DrugTest was the easiest to use and it showed a sensitivity ≥50% for all drug classes tested, with the specificity always above 90%. The sensitivity of DrugTest was below 90% for all drug classes, and 81% for cannabis, the most prevalent drug among impaired drivers in Belgium (3). Overall, compared to previous evaluations, some progress was noted in terms of sensitivity and specificity; however, more improvement is still required.

Acknowledgments
This project was funded by the European Commission as part of the DRUID project and the National Lottery of Belgium. We acknowledge the assistance provided by the staff of MSOC Ostend, De Sleutel in Wondelgem and UPSIE Ghent. Also we
would like to thank the manufacturers for providing the devices and training needed for this study.

Disclaimer
This article has been produced under the project Driving Under Influence of Drugs, Alcohol and Medicines (DRUID) financed by the European Community within the framework of the EU 6th Framework Program. This report reflects only the authors’ view. The European commission is not liable for any use that may be made of the information contained therein.

References