Multi-year interlaboratory exercises for the analysis of illicit drugs and metabolites in wastewater: development of a quality control system

Alexander L.N. van Nuijs[1], Foon Yin Lai[1], Frederic Been[1], Maria Jesus Andres-Costa[2], Leon Barron[3], Jose Antonio Baz-Lomba[4], Jean-Daniel Berset[5], Lisa Benaglia[6], Lubertus Bijlsma[7], Dan Burgard[8], Sara Castiglioni[9], Christophoros Christophoridis[10], Adrian Covaci[1], Pim de Voogt[11,12], Erik Emke[11], Despo Fatta-Kassinos[13], Jerker Fick[14], Felix Hernandez[7], Cobus Gerber[15], Iria González-Mariño[16], Roman Grabic[17], Teemu Gunnar[18], Kurunthachalam Kannan[19], Sara Karolak[20], Barbara Kasprzyk-Hordern[21], Zenon Kokot[22], Ivona Krizman-Matasic[23], Angela Li[24], Xiqing Li[25], Arndís S.C. Löve[26], Miren Lopez de Alda[27], Markus R. Meyer[28], Herbert Oberacher[29], Jake O'Brien[30], Jose Benito Quintana[16], Malcolm Reid[4], Serge Schneider[31], Susana Sadler Simoes[32], Nikolaos S. Thomaidis[33], Kevin Thomas[4,30], Viviane Yargeau[34], Christoph Ort[35]

[1] Toxicological Centre, University of Antwerp, Universiteitsplein 1, 2610 Antwerp, Belgium
[2] Environmental and Food Safety Research Group (SAMA-UV), Desertification Research Centre CIDE (CSIC-UV-GV), Av. Vicent Andrés Estellés s/n, Burjassot, Valencia, Spain
[3] Analytical & Environmental Sciences Division, Faculty of Life Sciences & Medicine, King's College London, Franklin Wilkins Building, 150 Stamford St., London SE1 9NH, United Kingdom
[4] Norwegian Institute for Water Research (NIVA), Gaustadalléen 21, 0349 Oslo, Norway
[5] Institute of Plant Sciences (IPS), University of Bern, Altenbergrain 21, 3013 Bern, Switzerland
[6] École des Sciences Criminelles, University of Lausanne, Avenue Forel 15, 1015 Lausanne, Switzerland
[7] Research Institute for Pesticides and Water, University Jaume I, Avda. Sos Baynat s/n, E-12071 Castellón, Spain
[8] Chemistry Department, University of Puget Sound, Tacoma, WA, 98416, USA
[9] IRCCS – Istituto di Ricerche Farmacologiche "Mario Negri", Department of Environmental Health Sciences, Via La Masa 19, 20156 Milan, Italy
[10] Environmental Pollution Control Laboratory, Aristotle University of Thessaloniki, 54124, Greece
[11] KWR Watercycle Research Institute, Chemical Water Quality and Health, P.O. Box 1072, 3430 BB Nieuwegein, The Netherlands
[12] Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands
[13] Nireas-International Water Research Center and Civil and Environmental Engineering Department, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus
[14] Department of Chemistry, Umeå Unicversity, 901 87 Umeå, Sweden
[15] School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, Australia, 5001
[16] Institute for Food Analysis and Research, University of Santiago de Compostela, Constantino Candeira S/N, 15782 Santiago de Compostela, Spain
[17] University of South Bohemia in Ceske Budejovice, Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Zatisi 728/II, CZ-389 25 Vodnany, Czech Republic
[18] Forensic Toxicology Unit, National Institute for Health and Welfare, P.O.Box 30, 00271 Helsinki, Finland
[19] Wadsworth Center, New York State Department of Health, and Department of Environmental Health Sciences, School of Public Health, State University of New York at Albany, Empire State Plaza, Albany, NY 12201-0509, USA
[20] Public Health and Environnement Laboratory, UMR 8079 Ecologie Systématique Evolution, Faculty of

Pharmacy, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, 92296 Châtenay-Malabry, France
[21] University of Bath, Department of Chemistry, Faculty of Science, Bath BA2 7AY, United Kingdom
[22] Department of Inorganic and Analytical Chemistry, Poznan University of Medical Sciences, 6 Grunwaldzka Street, 60-780 Poznan, Poland
[23] Division for Marine and Environmental Research, Rudjer Boskovic Institute, Bijenicka 54, Zagreb, 10000 Croatia
[24] Food Safety Laboratory, Health Sciences Authority, Singapore
[25] Laboratory for Earth Surface Processes, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China
[26] Department of Pharmacology and Toxicology, University of Iceland, Hofsvallagata 53, 107 Reykjavik, Iceland
[27] Water and Soil Quality Research Group, Department of Environmental Chemistry, Institute of Environmental Assessment and Water Research (IDAEA-CSIC), Jordi Girona 18-26, 08034 Barcelona, Spain
[28] Department of Experimental and Clinical Toxicology, Center for Molecular Signaling (PZMS), Saarland University, 66421 Homburg, Germany
[29] Institute of Legal Medicine and Core Facility Metabolomics, Medical University of Innsbruck, Muellerstrasse 44, 6020 Innsbruck, Austria
[30] Queensland Alliance for Environmental Health Sciences (QAEHS), University of Queensland, 39 Kessels Road Coopers Plains, Queensland 4108, Australia
[31] Laboratoire National de Santé, Service de toxicologie analytique et de chimie pharmaceutique, 1 rue Louis Rech, L-3055 Luxembourg
[32] National Institute of Legal Medicine and Forensic Sciences, South Branch, Rua Manuel Bento de Sousa n◦3, 1169-201 Lisbon, Portugal
[33] Laboratory of Analytical Chemistry, Department of Chemistry, National and kapodistrian of Athens, Panepistimiopolis Zografou, 15771 Athens, Greece
[34] Department of Chemical Engineering, McGill University, Montreal, Quebec, Canada, H3A0C5
[35] Eawag, Swiss Federal Institute of Aquatic Science and Technology. Urban Water Management. Überlandstrasse 133, 8600 Dübendorf, Switzerland

Corresponding author:

Prof. Dr. Alexander L.N. van Nuijs
Toxicological Centre, University of Antwerp
Universiteitsplein 1
2610 Antwerp, Belgium
e-mail: alexander.vannuijs@uantwerpen.be
tel: +32 (0)3 265 24 98

Abstract

This study presents the development of a worldwide inter-laboratory testing scheme for the analysis of seven illicit drug residues in different matrices (standard solutions, tap- and wastewater). By repeating this exercise for six years with participation of 37 laboratories from 25 countries, the testing scheme was substantially improved based on experiences gained across the years (e.g. matrix type, sample conditions, spiking levels). From the exercises, (pre-)analytical issues (e.g. pH adjustment, filtration) were revealed for some analytes which resulted in formulation of best-practice protocols, both for inter-laboratory setup and analytical procedures. The results illustrate the effectiveness of the inter-laboratory testing scheme in assessing laboratory performance in the framework of illicit drug analysis in wastewater. The exercise proved that measurements of laboratories were of high quality (> 80% satisfactory results for 6 out of 7 analytes) and that analytical follow-up is important to assist laboratories in improving robustness of wastewater-based epidemiology results.

Keywords

Illicit drugs; wastewater; inter-laboratory testing; wastewater-based epidemiology; quality assurance

1. Introduction

The measurement of the human excretion products of illicit drugs in influent wastewater has been recognized as an alternative and complementary approach for estimating the consumption of illicit drugs within communities, i.e. the catchment of wastewater treatment plants (WWTPs) [1-3]. The principle behind wastewater-based epidemiology (WBE) derives from the fact that parent compounds and/or their human metabolites (i.e., drug residues) are excreted in urine and faeces following illicit drug use and end up in urban sewer systems [3]. The ability of WBE to provide useful and timely information on temporal (daily, weekly, monthly, and annually) and spatial (within- and between-countries) variations in illicit drug consumption has been demonstrated [4-15]. The European Monitoring Centre for Drug and Drug Addiction (EMCDDA) has recently acknowledged the added value of WBE to socio-epidemiological methods, such as population surveys, seizure data and crime statistics, in generating useful and relevant data on population drug use [3].

With the aim to improve and optimize WBE, a Europe-wide collaboration was initiated in 2010. Seven European institutions – University of Antwerp (BE), Eawag (CH), University Jaume I (ES), Mario Negri Institute (IT), KWR Watercycle Research Institute (NL), Norwegian Institute for Water Research NIVA (NO), and University of Bath (UK) - established the research group SCORE (Sewage analysis CORe group Europe) [16]. The ultimate goals of SCORE are (a) to collaborate in the field of WBE to provide reproducible data; (b) to improve and harmonize the analytical procedures used in different laboratories to analyze drug residues in wastewater samples; and (c) to perform international studies comparing illicit drug consumption in communities across the world. To this end, SCORE has coordinated monitoring studies and exercises to assure the quality of reported data based on agreed best-practices tackling sampling, storage and analysis. Important results from this collaboration are multi-city studies demonstrating the usefulness of WBE on an international level to obtain the most recent data on illicit drug consumption [17-18].

In order to further optimize and fine-tune WBE, it is imperative to gain knowledge on the sources of uncertainties that are associated with the approach. In 2013, SCORE performed a thorough evaluation on the uncertainties of WBE using the best-practice protocols and data that were available from the comparative Europe-wide WBE research [19]. One of the cornerstones of WBE is to accurately quantify concentrations of drug residues in wastewater samples by means of reliable analytical procedures [20]. This requires fully validated analytical procedures before routine analysis can be initiated and participation in external quality control schemes is, where possible, highly recommended. External quality control through inter-laboratory exercises are based on the

distribution of the same test samples (in our case prepared by NIVA) to all participants. The latter analyse all test samples without any knowledge of the concentrations of target analytes and return their results to the coordinator of the exercise (in our case Eawag, who does not analyse test samples and does not know the nominal spike value until final compilation of results). The coordinator converts the submitted results into objective scores that reflect the performance of individual laboratories and the group. These scores can alert participants of unexpected problems and can result in actions to be taken [21].

SCORE initiated inter-laboratory exercises in 2011 in order to develop a quality control scheme for laboratories that analyze illicit drug residues in wastewater for WBE purposes. Since its debut, the testing scheme has been carried out annually with increasing participation of different laboratories, also extending the network outside Europe. The objectives of the presented interlaboratory exercise are (a) to illustrate the results of the six-year inter-laboratory testing scheme; (b) to evaluate advancements achieved over these years and to identify issues still to be resolved; (c) to formulate recommendations for future inter-laboratory exercises and (d) to propose a robust quality control system to improve the analytical performance of laboratories analyzing illicit drugs in wastewater.

2. Setup of the inter-laboratory exercises

2.1. Target analytes

A total of seven illicit drug residues were targeted in the inter-laboratory testing scheme. These included cocaine (COC), benzoylecgonine (BE, cocaine metabolite), 3,4-methylenedioxy-methamphetamine (MDMA), amphetamine (AMP), methamphetamine (METH), 11-nor-9-carboxy-tetrahydrocannabinol (THC-COOH, THC metabolite), and 6-monoacetylmorphine (6-MAM, heroin metabolite). These analytes are widely regarded as the main urinary biomarkers of the worldwide most consumed illicit drugs (COC, MDMA, AMP, METH, cannabis and heroin) and are the focus of most bioanalytical and WBE initiatives around the world [22]. Certified spiking solutions of each of the target analytes were supplied by Cerilliant Corporation (Round Rock, Texas, USA). All spiking solutions were supplied in sealed glass ampoules at 1 mg/mL in methanol.

2.2. Design of the exercises

The basis of the inter-laboratory testing scheme was to compare the performance of the analytical procedures employed by participating laboratories. Two separate modules were included to evaluate in each laboratory (a) the use of correct analytical reference standards and the performance of the

instrumental analysis (Module 1), and (b) the performance of entire analytical procedures applied to the analysis of wastewater, including sample preparation (Module 2).

For Module 1, a methanol solution containing the seven target analytes was used. For Module 2, samples of tap water and wastewater spiked with the seven analytes were employed. Participants were asked to use their own in-house developed and validated analytical procedures for the analysis of the samples. Replicate analysis of each sample was requested (n = 5 for Module 1 and n = 3 for Module 2). Commonly, sample pre-treatment consisted of filtration followed by solid-phase extraction for Module 2 samples. All laboratories employed liquid chromatography coupled to mass spectrometry using mass-labelled internal standards to perform detection and quantification of the analytes. More information on different techniques, including sample preparation procedures, used for this type of analyses can be found in Castiglioni et al. (2013) and Hernandez et al. (in press) [19-20].

Analyte stability in various matrices and conditions is a crucial aspect of any inter-laboratory exercise as it can substantially affect the outcomes of the analyses, particularly in the absence of certified reference material in target matrices. Stability of illicit drugs in wastewater has been the subject of numerous investigations, which were recently reviewed by McCall et al. (2016) [23]. Detailing the results from all these studies goes beyond the scope of the present paper, however, a brief overview regarding the analytes targeted in this inter-laboratory exercise is reported here. Both COC and BE have been shown to be stable in wastewater over multiple weeks when stored refrigerated (4 °C and, ideally, -20 °C), at low pH and in the dark. Similarly, MDMA, AMP and METH have been shown to be stable under similar conditions. THC-COOH and 6-MAM, on the other hand, have been shown to be very sensitive to temperature and, for THC-COOH, low pH.

2.3. Preparation of test samples

All test samples were prepared by the Norwegian Institute for Water Research (NIVA). Figure 1 and Table 1 give an overview of the type of test samples included in each year (2011-2016) and the nominal spiking levels used. The two modules together comprised three matrices (i.e., methanol, tap water and wastewater) spiked at different concentrations for each of the target analytes. Spiking concentrations for all matrices changed from year to year to avoid bias and ensure legitimate results. Certified spiking solutions (1 mg/mL in methanol) were diluted to prepare working solutions at 100 µg/mL or 10 µg/mL in methanol. The working solutions were then used to prepare different test samples.

The methanol solution (Module 1) containing the analytes was prepared from each of the 100 µg/mL working solutions. Aliquots (1 mL) of this methanol sample were then transferred to separate glass

vials and capped. Each vial was accurately weighed and stored at -20 °C ahead of shipment to the participants. Participants were asked to weigh the samples at arrival and to report deviations from the weight at preparation.

Spiked wastewater and tap water samples (Module 2) were prepared in a 20 L high-density polyethylene (HDPE) plastic container pre-washed with tap water and methanol. Twenty litres of cold tap water or fresh wastewater from VEAS WWTP in Oslo (Norway) were poured into the container, spiked with different volumes of the 10 µg/mL working standard solutions to obtain relevant concentrations (at ng/L range) and stirred for 2 h to homogenize the mixture. In 2012, one of the wastewater samples was used as it is; no spiking with target analytes occurred.

Samples from Module 2 were acidified to adjust the pH to 3.5 in 2012 and 2013. This pH adjustment was agreed upon by the organizers of the exercise as at that time it was assumed that acidification of samples was the best way to prevent degradation of the analytes [19]. In 2014-2016, no pH adjustment of the tap water was performed because of the new insight into the negative effect of low pH on the stability of THC-COOH in wastewater [23-24]. The changes in used matrices and pH conditions across the years of the inter-laboratory exercise were the result of experiences of previous years and of advancements made in the field of WBE.

Aliquots of at least 250 mL were placed in HDPE containers and stored at -20 °C before shipping to the participants. As real wastewater was used, and which likely contained unknown concentrations of the target analytes, it was not possible to use a genuine "blank" wastewater sample and nominal values could thus not be reported. Instead, a total value, comprising background concentrations (x) and the spiked level, was computed (Table 1).

2.4. Participants and sample shipping

The inter-laboratory exercises were organized by SCORE and were open to interested participants from any institution. In order to participate to the exercise, laboratories were required to register (without any payment) following an invitation sent out by SCORE or through the SCORE website [16]. Over the period between 2011 and 2016, a total of 37 laboratories from 25 countries participated in the exercises (for more details on participation in each year, see Table 1). Most of the participating laboratories (81%) were located in Europe, while the rest (19%) was spread over different continents (North-America, Asia and Oceania) (Figure 2). The participants located within the European Union received the test samples, shipped on ice, during the following 24-48 hours while for the remaining participants from the other continents the average transport time was 2-4 days. Temperature during shipment was not recorded, but participants were asked to not analyse samples if defrosted upon reception (responsibility if the participant).

2.5. Evaluation of results

Participating laboratories were required to report measured concentrations of the target analytes in each sample type provided. Results of individual replicates were submitted. Furthermore, participants had to clearly highlight when concentrations were not quantifiable (i.e., below limits of quantification) or when the analysis for a certain compound was not performed. Limits of quantification for each participant were estimated with a fixed protocol and compared to self-assessed limit of quantifications. It was established at a signal-to-noise ratio of 10 using the quantifier transition from chromatograms of samples spiked at the lowest validation level tested. The estimated limits of quantification were for all participating laboratories within the same order of magnitude and comparable to what was reported by each lab based on validation data. Since 2015, one spiking level was used to evaluate whether the analytical procedures of participants had limit of quantifications that are relevant in the context of WBE studies. If participants could not report values for this sample, they were notified that their analytical procedures did not reach relevant sensitivity.

First, the mean concentration (m) of replicates for each participant and for each sample type was calculated. Secondly, after testing for normality, a Grubbs' test was performed to identify outliers which were excluded from further analysis. From the remaining means, the group's mean [i.e., mean of means (M)] and the group's standard deviation (SD) were computed. To evaluate the performance of each participant (i), z-scores (z_i) for every analyte and sample type were calculated as follows:

$$z_i = \frac{m_i - M}{SD}$$

Following the ISO standard, a laboratory passed the inter-laboratory exercise when its |z| ≤ 2 [21, 25]. Participants with results that were identified as outliers (Grubb's test) or had |z|-values > 2 were individually notified about the deviation and were allowed to recheck their submitted values for inconsistencies or errors. Note that no detail (z_i, M) was supplied with the notification of the deviation in order to maintain impartiality. If these laboratories were able to supply a viable explanation (such as transcription errors), they were allowed to resubmit corrected results. If accepted, newly submitted values were used to compute updated values for m_i, M, SD and z_i.

The purpose of this iterative process lies in the goal of SCORE to advance and improve WBE. The inter-laboratory exercise was therefore used to assist laboratories in optimizing their analytical procedures and improve the overall performance.

3. Results and Discussion

3.1. Assigned value: group's mean vs. nominal concentration

The z-score was calculated relative to the group's mean (M). The main reasons for using M instead of the nominal concentration (i.e. spiking levels) as reference in the context of this inter-laboratory exercise are [21, 25]:

(i) Multiple scientific evaluations repeatedly revealed that spiking concentration levels did not necessarily display sufficient reliability to be used as an assigned value to calculate z-scores;

(ii) For wastewater samples, the use of spiking levels as assigned value is out of the question because of the presence of unknown concentrations of the analytes (no nominal values exist);

(iii) There is a sufficient number of laboratories that participated in the exercises along the years (Table 1);

(iv) Certified reference materials (CRMs) for analyzing illicit drugs in water samples are not available;

(v) No recognised reference laboratories for this type of analysis exist;

(vi) The chosen approach was agreed by the participants as they were all informed on the calculation and evaluation procedures applied.

Figure 3 shows the deviation of the group's mean (M) from the nominal concentration (spiking level) for the methanol and tap water test samples. For the wastewater samples included in the exercises from 2012-2014, it is impossible to generate any meaningful plot because of the unknown background concentrations of the analytes present in this matrix.

The results showed that the deviation of the group's mean (M) from the nominal concentration was mostly < 25%, which was regarded by SCORE as an acceptable variability. The deviation for the matrix-free samples (i.e., methanol solvent) was mostly well below this 25% limit and suggested that in all laboratories, the reference standards (both native and isotope-labelled) used and the instrumental analysis (e.g. calibration and instrumental parameters) did not lead to substantial bias in the analysis of the target analytes, except for 6-MAM. However, in the presence of matrix, deviations of more than 25% occurred more often, in particular for 6-MAM and THC-COOH. Concentrations of 6-MAM were systematically underreported, for both the standard solution and tap water samples. In some occasions, the deviation amounted up to 60%. This systematic underestimation of 6-MAM could be due to: (i) inaccuracies during the preparation and spiking of the test samples (e.g. preparation and dilution of stock solution); (ii) stability issues of this analyte during preparation of the test samples and during storage and sample handling; (iii) issues with the analytical procedures applied by the laboratories.

The analysis of THC-COOH in the methanol samples gave acceptable results (deviation <25% and no systematic error), while deviations of up to 90% were observed in tap water samples in 2013 and 2014. It is important to highlight that tap water samples were acidified in 2013 and, in the following year, sample acidification before filtration was still performed by multiple participants. These were later shown to have a negative impact on the measured concentrations of THC-COOH because of adsorption issues [23-24, 26]. Acidification may be the cause of the high variability observed for this analyte, but this is clearly not the whole picture. In fact, Causanilles et al. (2017) demonstrated that different (combinations of) parameters (pH, filtration, sorption) can have an influence on the analysis of THC-COOH in wastewater [26].

For COC, all samples across the different years showed deviations <25%, except for the three tap water samples in 2015. The nature of this systematic deviation (only one year) indicates the error likely occurred in the preparation of these test samples.

3.2. Influence of different matrices and concentration levels on the group's variability

The influence of the different matrix types on the performance of participating laboratories was assessed through analysis of the datasets from all years. Figures 4 and 5 illustrate the influence of the three matrices on the relative standard deviation (RSD) of the group. Overall, a lower RSD for the methanol samples compared to the waste- and tap water samples was observed (Wilcoxon rank sum test p-value < α = 0.05). This observation was not surprising considering that concentrations of the standard solution samples were in the µg/L range while in tap water and wastewater, samples concentrations were in the ng/L range. Furthermore, analysis of the methanol solution samples did not require any substantial sample preparation (i.e., direct injection with/without further dilution) compared to waste- and tap water samples, which required pre-concentration. A significant difference between the RSDs for tap water and wastewater samples was observed (Wilcox rank sum test p-value = 0.01, α = 0.05). For THC-COOH, high RSDs were observed for tap water and wastewater samples compared to the other analytes. Likewise, in the methanol solution, high RSDs were observed on several occasions (Figure 4). These findings further suggest that there are some issues with the analysis of this particular compound in water samples, as discussed earlier (Figure 3).

The difference in RSDs between tap and wastewater samples was further investigated using ANOVA (after log transforming the data to correct for deviation from normality and heteroscedasticity). Statistical analysis revealed that the spiking level showed the most significant influence on the group's RSD ($F(1,98) = 121.5$, $p < 0.0001$), followed by the matrix type ($F(1,98) = 10.9$, $p < 0.001$) and the compound under analysis ($F(6,98) = 3.0$, $p < 0.01$). Because the matrix type was not the most influential parameter, the use of spiked tap water samples was deemed adequate for the purposes of the present inter-laboratory exercise. In fact, when using wastewater samples, (a) differences in

matrix effects occur between locations and (b) background concentrations of the analytes in wastewater are unknown and uncontrollable. As a result, it was not considered possible to use 'representative' wastewater for the purpose of this inter-laboratory exercise. Furthermore, by using tap water, labour and logistic costs linked to the preparation and distribution of additional samples to the participants could be reduced significantly. Issues related to the biodegradation and sorption of target analytes in wastewater during shipment could also be reduced. Furthermore, our study, including data over a six-year period, provides unique insights into how the molecular properties of the analytes, concentration levels and matrix type affect laboratory performance in the context of (waste)water analysis. The information and experience gained could hence be useful for other inter-laboratory exercises confronted with similar matrices.

3.3. Performance of laboratories

The evaluation of the results obtained by all laboratories discussed hereafter is based on the performances with the spiked tap water samples, as this matrix was shown to be appropriate (see section 3.2) and because of the issues with wastewater samples mentioned earlier (i.e., unknown background concentrations and potential stability issues). Figure 6 provides an overview of the proportion of satisfactory results per analyte type in the period of 2013-2016. A satisfactory result is regarded as a $|z|$-value ≤ 2 [21, 25]. Grubb's outliers, non-detects (reported as below limit of quantification) and $|z|$-values > 2 are regarded as unsatisfactory. In the supporting information, detailed results for each laboratory over the different years are shown. The plots give an overview of the distribution of the z-scores of the group for the different years, matrices and spiking levels and detailed plots for results of the individual laboratories (including intra-laboratory variation).

In general, for BE, COC, MDMA, and AMP, the group's performances were acceptable, with > 90% of satisfactory results. For METH and 6-MAM, the satisfactory result were around 80% in 2013. This can be linked to the fact that 3 out of 15 (METH) and 3 out of 10 (6-MAM) participants did not detect the analytes in the test samples. In 2014-2016, acceptable results for these two analytes were obtained, probably due to the higher concentration levels and improved performance of the analytical procedures of the participants. The unsatisfactory results obtained for THC-COOH analysis over years have drawn the attention of SCORE and triggered a further investigation of the effect that different pre-analytical steps (filtration and pH adjustment) have on the accuracy the analysis of this compound in wastewater [26].

It is important to mention that the aim of SCORE is to improve the reliability of WBE studies. Therefore, support was provided to laboratories that showed unsatisfactory results by means of short-term visits of a SCORE member and/or optimization of the analytical procedures (assistance with sample preparation and method validation). In most cases, this resulted in positive outcomes

for these laboratories in following exercises. This highlighted the need for follow-up of inter-laboratory exercises combined with a continuous support to all participants.

The z-scores regarding different concentrations of each analyte were visualised in scatter biplots (i.e., Youden plots, Figure 7) to assess the sources of variability among the participating laboratories. Inter-laboratory variation predominates if results were clustered in the upper right and lower left (= white) quadrants, while intra-laboratory variation predominates if results are clustered in the upper left and lower right (= grey) quadrants [25]. Furthermore, the distances of the plotted point relative to the 45-degree reference line and to the (0, 0) point (i.e. the Manhattan median) are both useful for the interpretation of inter-laboratory data. Points that lie close to the 45-degree reference line but far from the Manhattan median indicate a systematic error. Points that lie far from the reference line suggest large random errors. The majority of the participating laboratories was found within the white quadrants (Figure 7), meaning that inter-laboratory variability was predominant over the intra-laboratory variability for all seven analytes. Only a few laboratories were occasionally outside of the $|z|$-values > 2 boundaries. For the latter, this implies large total errors, which were mainly systematic, as results were close to the 45-degree reference line but distant from the origin. Moreover, it should be noted that no recurrent erroneous results were observed, i.e., there were no laboratories with anomalous results for a certain analyte reported across different years. This supports the hypothesis that the observed errors were rather incidental and/or that these laboratories had improved their analytical procedures.

3.4. Sources of variations and recommendations

The six-year data from inter-laboratory exercises for the analysis of illicit drug residues in water samples revealed variations linked to its setup and allowed to provide recommendations to improve future exercises. First, this study shows that the group's mean should be used to evaluate performance of laboratories rather than the nominal (spiked) value. However, it is important that nominal values should always be considered to exclude pre-analytical issues, as demonstrated for THC-COOH. This observation triggered further investigations and recommendations to improve the WBE approach to estimate cannabis use [26]. Second, since concentration levels were found to be the main factor influencing performances (Figure 4, see section 3.2), spiking levels should be chosen carefully, and reflecting concentrations expected in real samples. Particularly, for the methanol standard samples, the use of different concentrations (e.g. Youden couple) instead of a single (high) level, as we did, will be useful to improve the assessment of laboratory performances. Third, it is important to prepare and transport test samples in the most optimal way in order to avoid stability and adsorption problems. The issues observed with 6-MAM and THC-COOH when samples were

404 acidified (see section 3.1) are a good example and highlight the need to consider other preservatives
405 (e.g., sodium metabisulphite ($Na_2S_2O_5$) or sodium azide (NaN_3)) to ensure analyte stability during
406 transport and storage [27-28]. Furthermore, future inter-laboratory exercises should include an extra
407 analysis of the test samples by the preparing laboratory directly after preparation of the test samples
408 before freezing and shipment. This will improve understanding of the differences between the
409 nominal spike and the assigned value.

410 Based on the experiences acquired from these six rounds of inter-laboratory exercises,
411 recommendations related to analytical procedures used by individual laboratories for measuring
412 illicit drugs and metabolites in wastewater can be formulated. Laboratories can freely choose their
413 preferred sample preparation procedure and detection/quantification technique, but we strongly
414 suggest that the methods comply with the following features. First, mass-labeled internal standards
415 should be used for each analyte and spiked in samples before any filtration step. Second, pH
416 adjustment - when needed - has to be conducted after internal standard spiking and/or filtration.
417 This is particularly relevant for the analysis of THC-COOH in wastewater [26]. Third, freeze-thaw
418 cycles of the samples should be minimized. Fourth, in-house quality control samples (e.g. spiked tap
419 water or wastewater) should be prepared and analysed with each sample batch. Furthermore,
420 centrifugation instead of filtration can be an alternative way to avoid the blockage and clogging of
421 solid-phase extraction cartridges with particulates present in the wastewater.
422
423 **4. Conclusions**
424 This study presents, for the first time, the results of an inter-laboratory testing scheme for the
425 analysis of illicit drugs and metabolites in wastewater. By repeating this exercise for six years, we
426 were able to improve the set-up of the testing scheme substantially, based on experiences gained
427 over the years (e.g. matrix to be used, sample parameters, spiking levels) and to establish a reliable
428 quality control system. The existence of such system is important to ensure high-quality data of WBE
429 monitoring studies that can be used by stakeholders to obtain the most recent data on spatial and
430 geographical trends in illicit drug use on a national and international scale.
431 The results of the exercise highlighted the importance of using the group's mean rather than the
432 nominal value as the assigned value, in particular due to the lack of certified reference materials for
433 testing illicit drugs in wastewater. An investigation of the RSD associated with reported results
434 showed that the most influential parameter was the spiking level, not the instrument (method) used
435 or the type of matrix (i.e., tap or wastewater). Consequently, tap water was chosen for future
436 exercises as it presents various advantages. Specifically, it allows to control spiking levels more easily,
437 which is not possible with wastewater as unknown background concentrations exist. In fact,

substantial variations in composition and analyte concentrations occur, even within wastewater collected from a unique location.

Regarding laboratories performances, the results from the inter-laboratory exercise show that these were generally satisfactory for COC, BE, MDMA, AMP and METH. An improvement was observed over the years and, in its latest round in 2016, more than 90% of the participating laboratories reported results $|z|$-value ≤ 2. In the case of 6-MAM and THC-COOH, results from the exercise showed that important pre-analytical issues still exist, and that sample pH has an important influence on the stability of the latter analytes. Whilst these issues still need to be solved, it is important to notice that none of the participating laboratories repeatedly (i.e., systematically) reported erroneous results for the same analyte across multiple years, emphasising the improvements in analytical performances which took place over the years.

The results illustrate the effectiveness of the inter-laboratory testing scheme in assessing and improving laboratory performance in the framework of illicit drug analysis in wastewater. The exercise proved that measurements of individual laboratories were of high quality and that analytical follow-up is important in order to assist laboratories in improving the robustness and accuracy of WBE results. The set-up and procedures used in this exercise for the measurement of illicit drugs in wastewater and experiences gained during the six-year period are of importance for the development of other quality control systems dealing with the measurement of pharmaceuticals, personal care products and other contaminants in aqueous matrices.

Wastewater-based epidemiology has gained importance, as numerous national and international organisations rely on its measurements to improve quantification of illicit drug use. Consequently, additional efforts will be needed in future to ensure the impeccable quality of reported results and tackle the existing and upcoming challenges. In particular, improving analytical performances for important compounds such as 6-MAM and THC-COOH and, at the same time, adapting protocols to integrate an ever growing number of relevant substances (e.g., new psychoactive substances) are among the main challenges that laboratories will face in future.

Acknowledgements

This article is based upon work from COST Action ES1307 supported by COST (European Cooperation in Science and Technology). We wish to acknowledge EMCDDA and Yeonsuk Ryu for support in the organization of the scheme and assistance in the preparation of the test samples, respectively. The following funding sources are acknowledged: the Research Foundation – Flanders (FWO), the Spanish Ministry of Economy, Industry and Competitiveness, the Generalitat Valenciana, *Xunta de Galicia,* Stavros Niarchos Foundation, Office for Combating Narcotic Drug Abuse of the Government of the Republic of Croatia, EU FP7 project SOLUTIONS (603437), the Government of Catalonia, the Natural

473 Sciences and Engineering Research Council of Canada (NSERC), Ministry of Education, Youth and
474 Sports of the Czech Republic (projects CENAKVA and CENAKVA II), EU Marie Skłodowska-Curie
475 Fellowship (APOLLO 749845) and the Swiss National Science Foundation (SNSF, P2LAP2_164892).
476 The following persons are acknowledged for help in sample analysis: Marijan Ahel, Evroula Hapeshi,
477 Popi Karaolia, Esther López-García, Nicola Mastroianni, Cristina Postigo, Inés Racamonde, Rosario
478 Rodil, Isaac Rodríguez, Tania Rodríguez-Álvarez, Ivan Senta, , and Senka Terzic, .

References

1. van Nuijs ALN, Castiglioni S, Tarcomnicu I, Postigo C, Lopez de Alda M, Neels H, Zuccato E, Barcelo D, Covaci A. Illicit drug consumption estimations derived from wastewater analysis: a critical review. Sci Total Environ. 2011a;409:3564-77

2. Castiglioni S, Thomas KV, Kasprzyk-Hordern B, Vandam L, Griffiths P. Testing wastewater to detect illicit drugs: state of the art, potential and research needs. Sci Total Environ. 2014;487:613-20

3. European Monitoring Centre for Drugs and Drug Addiction. Assessing illicit drugs in wastewater: advances in wastewater-based drug epidemiology, Insights 22, Publications Office of the European Union, 2016, Luxembourg

4. Harman C, Reid M, Thomas KV. In situ calibration of a passive sampling device for selected illicit drugs and their metabolites in wastewater, and subsequent year-long assessment of community drug usage. Environ Sci Technol. 2011;45:5676-82

5. van Nuijs ALN, Mougel JF, Tarcomnicu I, Bervoets L, Blust R, Jorens PG, Neels H, Covaci A. Sewage epidemiology--a real-time approach to estimate the consumption of illicit drugs in Brussels, Belgium. Environ Int. 2011b;37:612-21

6. Nefau T, Karolak S, Castillo L, Boireau V, Levi Y. Presence of illicit drugs and metabolites in influents and effluents of 25 sewage water treatment plants and map of drug consumption in France. Sci Total Environ. 2013;461-462:712-22

7. Mackuľak T, Skubák J, Grabic R, Ryba J, Birošová L, Fedorova G, Spalková V, Bodík I. National study of illicit drug use in Slovakia based on wastewater analysis. Sci Total Environ. 2014;494-495:158-65

8. Ort C, Eppler JM, Scheidegger A, Rieckermann J, Kinzig M, Sörgel F. Challenges of surveying wastewater drug loads of small populations and generalizable aspects on optimizing monitoring design. Addiction. 2014;109:472-81

9. Ostman M, Fick J, Näsström E, Lindberg RH. A snapshot of illicit drug use in Sweden acquired through sewage water analysis. Sci Total Environ. 2014;472:862-71.

10. Been F, Bijlsma L, Benaglia L, Berset JD, Botero-Coy AM, Castiglioni S, Kraus L, Zobel F, Schaub MP, Bücheli A, Hernández F, Delémont O, Esseiva P, Ort C. Assessing geographical differences in illicit drug consumption--A comparison of results from epidemiological and wastewater data in Germany and Switzerland. Drug Alcohol Depend. 2016;161:189-99

11. Kankaanpää A, Ariniemi K, Heinonen M, Kuoppasalmi K, Gunnar T. Current trends in Finnish drug abuse: Wastewater based epidemiology combined with other national indicators. Sci Total Environ. 2016;568:864-74

12. Krizman I, Senta I, Ahel M, Terzic S. Wastewater-based assessment of regional and temporal consumption patterns of illicit drugs and therapeutic opioids in Croatia. Sci Total Environ. 2016;566-567:454-62.

13. Lai FY, O'Brien JW, Thai PK, Hall W, Chan G, Bruno R, Ort C, Prichard J, Carter S, Anuj S, Kirkbride KP, Gartner C, Humphries M, Mueller JF. Cocaine, MDMA and methamphetamine residues in wastewater: Consumption trends (2009-2015) in South East Queensland, Australia. Sci Total Environ. 2016;568:803-9

14. Zuccato E, Castiglioni S, Senta I, Borsotti A, Genetti B, Andreotti A, Pieretti G, Serpelloni G. Population surveys compared with wastewater analysis for monitoring illicit drug consumption in Italy in 2010-2014. Drug Alcohol Depend. 2016;161:178-88.

15. Mastroianni N, López-García E, Postigo C, Barceló D, López de Alda M. Five-year monitoring of 19 illicit and legal substances of abuse at the inlet of a wastewater treatment plant in Barcelona (NE Spain) and estimation of drug consumption patterns and trends. Sci Total Environ. 2017;609:916-926.

16. SCORE (2010) Sewage analysis CORE group Europe. URL: http://score-cost.eu. Accessed: 2017-09-07. (Archived by WebCite® at http://www.webcitation.org/6tIO1NrbC)

17. Thomas KV, Bijlsma L, Castiglioni S, Covaci A, Emke E, Grabic R, Hernández F, Karolak S, Kasprzyk-Hordern B, Lindberg RH, Lopez de Alda M, Meierjohann A, Ort C, Pico Y, Quintana JB, Reid M, Rieckermann J, Terzic S, van Nuijs ALN, de Voogt P. Comparing illicit drug use in 19 European cities through sewage analysis. Sci Total Environ. 2012;432:432-9

18. Ort C, van Nuijs ALN, Berset JD, Bijlsma L, Castiglioni S, Covaci A, de Voogt P, Emke E, Fatta-Kassinos D, Griffiths P, Hernández F, González-Mariño I, Grabic R, Kasprzyk-Hordern B, Mastroianni N, Meierjohann A, Nefau T, Ostman M, Pico Y, Racamonde I, Reid M, Slobodnik J, Terzic S, Thomaidis N, Thomas KV. Spatial differences and temporal changes in illicit drug use in Europe quantified by wastewater analysis. Addiction. 2014;109:1338-52

19. Castiglioni S, Bijlsma L, Covaci A, Emke E, Hernández F, Reid M, Ort C, Thomas KV, van Nuijs ALN, de Voogt P, Zuccato E. Evaluation of uncertainties associated with the determination of community drug use through the measurement of sewage drug biomarkers. Environ Sci Technol. 2013;47:1452-60

20. Hernández F, Castiglioni S, Covaci A, de Voogt P, Emke E, Kasprzyk-Hordern B, Ort C, Reid M, Sancho JV, Thomas KV, van Nuijs ALN, Zuccato E, Bijlsma L. Mass spectrometric strategies for the investigation of biomarkers of illicit drug use in wastewater. Mass Spectrom Rev. in press (doi: 10.1002/mas.21525)

21. Thompson M, Ellison SL, Wood R. The international harmonized protocol for the proficiency testing of analytical chemistry laboratories. Pure Appl Chem. 2006; 78, 145-196

22. Baselt R. Disposition of toxic drugs and chemicals in man. 11th edition, Biomedical Publications, Seal Beach, CA, 2017, ISBN 978-0-692-77499-1

23. McCall AK, Bade R, Kinyua J, Lai FY, Thai PK, Covaci A, Bijlsma L, van Nuijs ALN, Ort C. Critical review on the stability of illicit drugs in sewers and wastewater samples. Water Res. 2016;88:933-47

24. Senta I, Krizman I, Ahel M, Terzic S. Assessment of stability of drug biomarkers in municipal wastewater as a factor influencing the estimation of drug consumption using sewage epidemiology. Sci Total Environ. 2014;487:659-65

25. ISO13528:2015(E). Statistical methods for use in proficiency testing by interlaboratory comparisons, ISO, 2015, Geneva, Switzerland

26. Causanilles A, Baz-Lomba JA, Burgard DA, Emke E, Gonzalez-Marino I, Krizman-Matasic I, Li A, Love ASC, McCall AK, Montes R, van Nuijs ALN, Ort C, Quintana JB, Senta I, Terzic S, Hernandez F, de

Voogt P, Bijlsma L. Improving wastewater-based epidemiology to estimate cannabis use: focus on the initial aspects of the analytical procedure. Anal. Chim. Acta 2017;988:27-33.

27. González-Mariño I, Quintana JB, Rodríguez I, Cela R. Determination of drugs of abuse in water by solid-phase extraction, derivatisation and gas chromatography-ion trap-tandem mass spectrometry. J Chromatogr A. 2010;1217:1748-60

28. Chen C, Kostakis C, Irvine RJ, Felgate PD, White JM. Evaluation of pre-analysis loss of dependent drugs in wastewater: stability and binding assessments. Drug Test Anal. 2013;5:716-21.

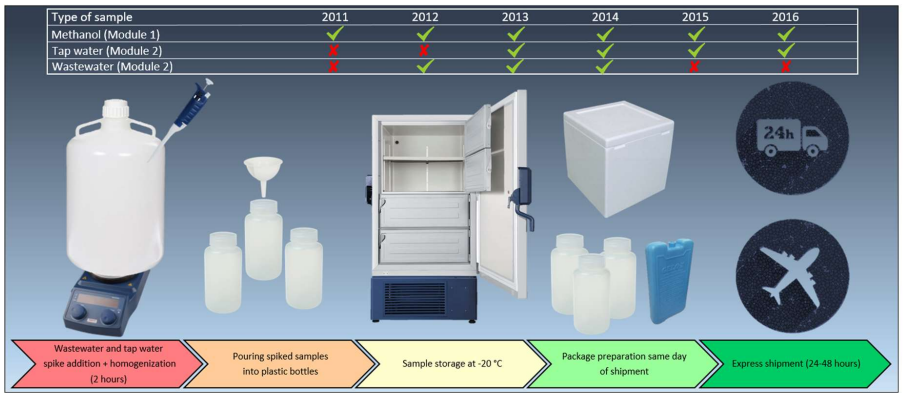

Figure 1. Inter-laboratory overview and scheme of the sample preparation and shipment for Module 2.

Figure 2. Map with location of the participants of the inter-laboratory exercises

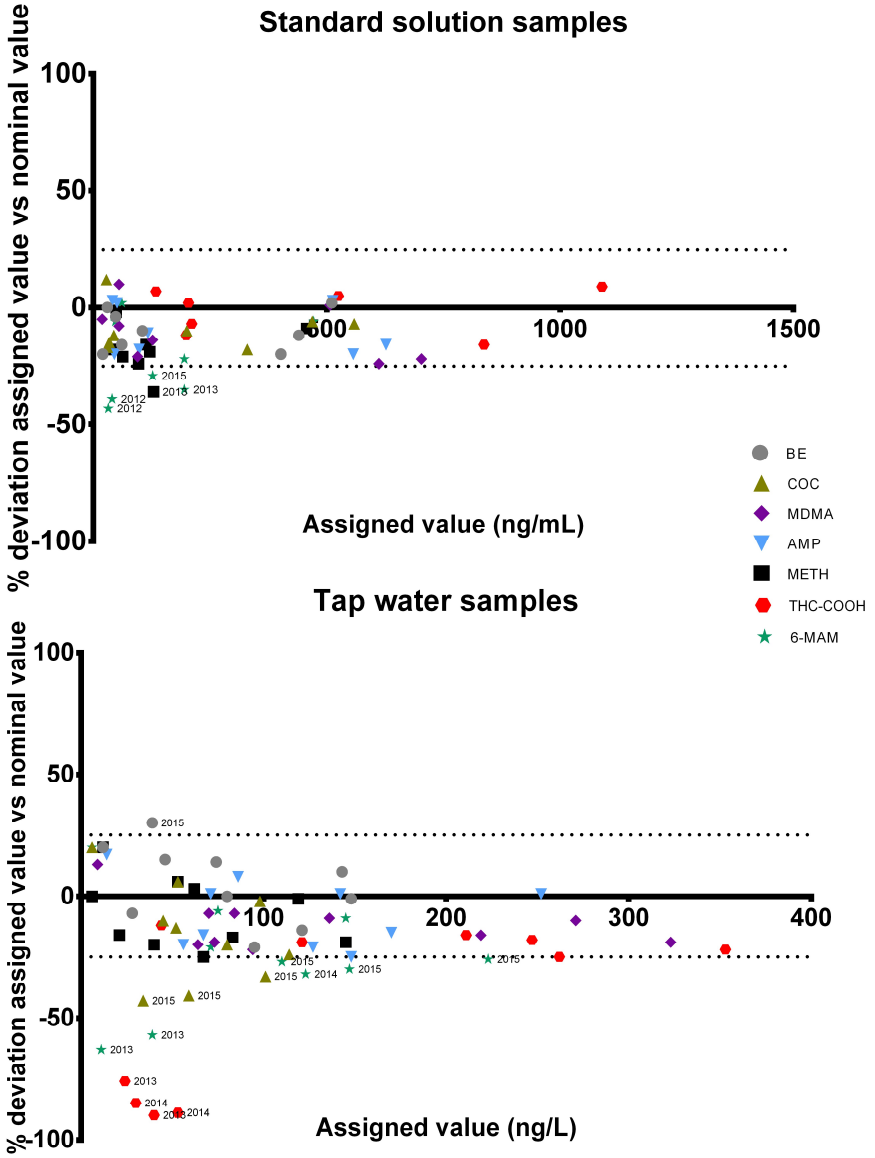

Figure 3. Deviation of the assigned value (= group's mean) from the nominal value (= spiking level) for the standard solution (top) and the tap water samples (bottom) in relation to the assigned value for the seven analytes. The dotted line represents 25% deviation. Entries with deviations > 25% are marked with the year of the inter-laboratory exercise.

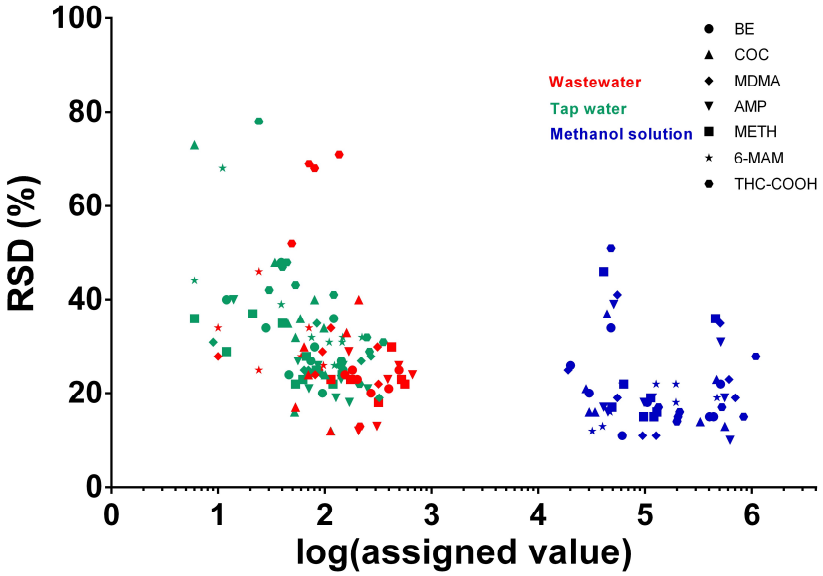

Figure 4. Relative standard deviation of the group in relation to the assigned value M (logarithmic scale) for the three matrices [standard solution (blue), tap water (green) and wastewater (red)] and seven analytes. All years (2011-2016) included.

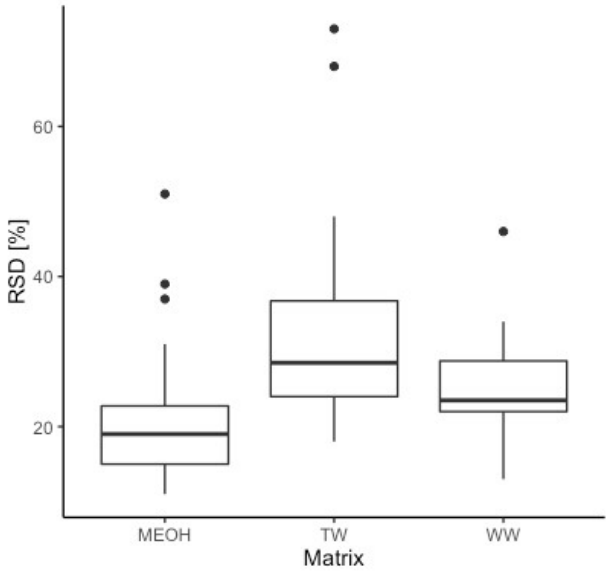

Figure 5. Boxplot showing the difference in the group's RSD for the three different matrices (MEOH = standard solution; TW = tap water; WW = wastewater) in 2013 and 2014 for all analytes.

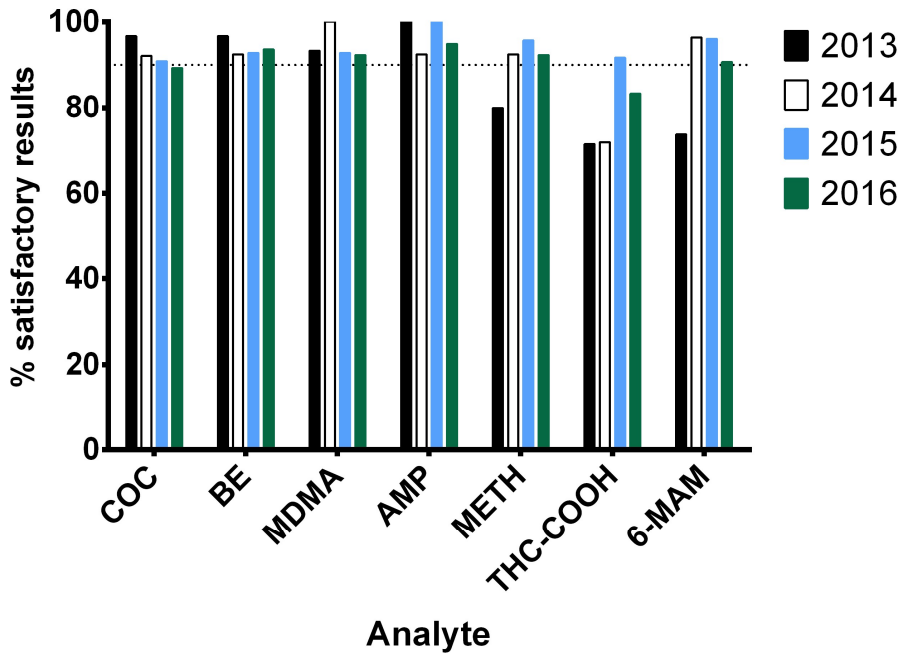

Figure 6. Percentage of participants with satisfactory results ($|z| \leq 2$) for tap water samples spiked with seven analytes. The dotted line represents 90% satisfactory level.

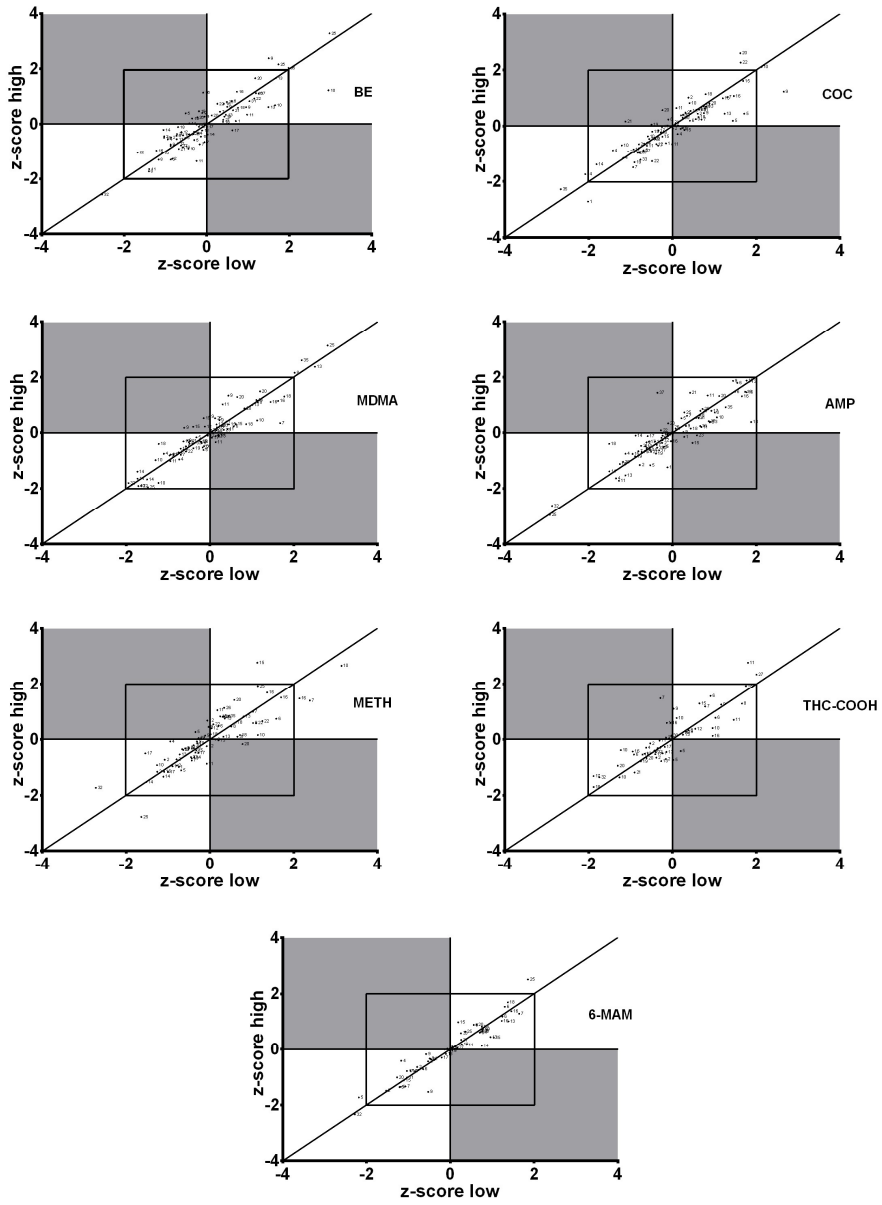

Figure 7. Youden plots with z-scores of the low concentration value (x-axis) and the z-scores of the high concentration value (y-axis) for the seven analytes in tap water across the years. Each participant is presented by a unique number. The inner rectangle captures satisfactory z-scores.