

# GC-C-IRMS in routine doping control practice: 3 years of drug testing data, quality control and evolution of the method

Michael Polet · Peter Van Eenoo

Received: 11 September 2014 / Revised: 24 November 2014 / Accepted: 27 November 2014 / Published online: 12 December 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** In order to detect the misuse of endogenous anabolic steroids, doping control laboratories require methods that allow differentiation between endogenous steroids and their synthetic copies. Gas chromatography combustion isotope ratio mass spectrometry (GC-C-IRMS) is capable of measuring the carbon isotope ratio of urinary steroids and this allows differentiation between both. GC-C-IRMS and its application to doping control has evolved a lot during the last decade and so have the World Anti-Doping Agency (WADA) technical documents that describe how GC-C-IRMS should be applied. Especially the WADA technical document of 2014 introduced a number of obligatory quality controls and a fixed methodology that should be used by all the doping control laboratories. This document imposed more uniform methods between the laboratories in order to decrease the interlaboratory standard deviation and acquire similar results for the analysis of the same urine samples. In this paper, 3 years of drug testing data of our GC-C-IRMS method in routine doping control practice is described, with an emphasis on the new WADA technical document and its implementation. Useful data for other doping control laboratories is presented focussing on general method setup, quality control and data collected from routine samples. The paper concentrates on how IRMS results shift or remain similar by switching to the 2014 WADA technical document and gives insight in a straightforward approach to calculate the measurement uncertainty.

**Keywords** GC-C-IRMS · Doping control · Steroids · Quality control

## Introduction

The administration of synthetic steroids is one of the most important issues facing sports. Synthetic copies of endogenous steroids are pharmacologically and chemically identical to their endogenous analogs, but there are small differences in the carbon isotope ratio (CIR). Natural variations in the abundance of  $^{13}\text{C}$  reflect the passage through biological and/or chemical processes in which transformations are accompanied by isotopic fractionation that results in a slight depletion or enrichment of  $^{13}\text{C}$  due to the difference in mass. CIRs are expressed as  $\delta^{13}\text{C}$  values against the international standard Vienna Pee Belemnite (VPDB) [1]. To measure the CIR of a compound, doping control laboratories use gas chromatography combustion isotope ratio mass spectrometry (GC-C-IRMS) [2–14]. Many laboratories derivatize their compounds before analysis. The CIR of the acetylated compounds needs to be corrected for the extra carbon atoms that have been incorporated into the steroidal skeletal. Therefore, the measured  $\delta^{13}\text{C}$  value has to be corrected using the following formula:

$$\delta^{13}\text{C}_{\text{steroid}} = \frac{(n_{\text{sa}}\delta^{13}\text{C}_{\text{sa}}) - (n_{\text{a}}\delta^{13}\text{C}_{\text{acorr}})}{n_{\text{steroid}}}$$

Here, ‘n’ is the number of moles of carbon, ‘sa’ stands for steroid acetate and ‘a’ for the acetate derivative group.  $\delta^{13}\text{C}_{\text{acorr}}$  is indirectly calculated by measuring the  $\delta^{13}\text{C}$  value of the reference material of each relevant analyte in the free and derivatized fractions.

Published in the topical collection on *Hormone and Veterinary Drug Residue Analysis* with guest editors Siska Croubels, Els Daeseleire, Sarah De Saeger, Peter Van Eenoo, and Lynn Vanhaecke.

M. Polet (✉) · P. Van Eenoo  
Department of Clinical Chemistry, Microbiology and Immunology,  
Doping Control Laboratory, Ghent University, Technologiepark 30  
B, 9052 Zwijnaarde, Belgium  
e-mail: michael.polet@ugent.be

GC-C-IRMS remains very laborious and expensive because one can only determine the CIR of a pure compound [15]. To minimize the amount of samples that needs to be analysed on GC-C-IRMS and maximize the detection of doping violations, all urine samples that enter a doping control laboratory are first screened by a fast gas chromatography triple quadrupole mass spectrometer (GC-MS/MS) method that quantifies all important endogenous steroids [16]. Afterwards, samples with suspicious endogenous steroid concentrations and/or ratios are forwarded to IRMS according to WADA regulations [17]. The testosterone/epitestosterone (T/E) ratio is the most important parameter for detecting T or T prohormone administration and its threshold is currently set at 4 by WADA as most individuals have values that vary between 0.1 and 3 [18, 19].

After administration of a synthetic anabolic steroid, the CIR of that steroid and its metabolites will exhibit depleted CIR in relation to an endogenous reference compound (ERC). ERCs are steroids that are not involved in the pathway of anabolic steroids and therefore their CIR will not be affected by application of an exogenous anabolic steroid. Pregnanediol (PD, 5 $\beta$ -pregnane-3 $\alpha$ ,20 $\alpha$ -diol) is used as primary ERC, but others such as 5 $\alpha$ -androst-16-en-3 $\alpha$ -ol, 11 $\beta$ -hydroxyandrosterone and 11-oxo-etiocholanolone (11-oxoEt, 5 $\beta$ -androstane-3 $\alpha$ -ol-11,17-dione) are also routinely used. ERCs normalize GC-C-IRMS results in relation to the diet of the individual athlete and WADA has set a minimum threshold of 3‰ difference between ERC and target compound (TC) in order to prove an adverse analytical finding by synthetic steroid abuse [17]. TCs are steroids that can be administered as doping substance or their metabolites (for example androsterone (A), etiocholanolone (Et), 5 $\alpha$ -androstane-3 $\alpha$ ,17 $\beta$ -diol ( $\alpha\alpha\beta$ -diol), 5 $\beta$ -androstane-3 $\alpha$ ,17 $\beta$ -diol ( $\beta\alpha\beta$ -diol),...).

The main goal of this paper was to provide useful data for other doping control laboratories and illustrate how we deal with GC-C-IRMS measurements, with an emphasis on the new WADA technical document and its implementation [17]. In an extensive method as GC-C-IRMS, localizing a malfunction can obviously be very time consuming. The different quality controls incorporated in the GC-C-IRMS method are described and an elucidation is given on how these can help identifying the source of deviating  $\delta^{13}\text{C}$  values.

## Materials and methods

### Reagents and chemicals

All used reagents and chemicals are mentioned in previous papers [20, 21].

## GC-C-IRMS confirmation procedure and its evolution

All samples were processed according to our routine IRMS confirmation procedure. The first samples were analysed in February 2011. Since then, the method has been adapted on several occasions in order to improve quality control, simplify the sample preparation, expand the number of TCs and ERCs and to comply with WADA requirements (Table 1) [17]. The original method started with aliquotation of a system blank, quality control negative (QC Neg), QC positive (QC Pos) and the routine samples, followed by solid phase extraction (SPE), hydrolysis, liquid extraction (LLE) with diethyl ether, acetylation and a high-performance liquid chromatography fraction collection (HPLC-FC). Four TCs (androsterone acetate (A-Ac), etiocholanolone acetate (Et-Ac), 5 $\alpha$ -androstane-3 $\alpha$ ,17 $\beta$ -diacetate ( $\alpha\alpha\beta$ -Ac<sub>2</sub>), 5 $\beta$ -androstane-3 $\alpha$ ,17 $\beta$ -diacetate ( $\beta\alpha\beta$ -Ac<sub>2</sub>)) and one ERC (5 $\beta$ -pregnane-3 $\alpha$ ,20 $\alpha$ -diacetate, PD-Ac<sub>2</sub>) were collected, analysed on GC-MS to check for peak impurities and finally the CIRs were determined on GC-C-IRMS by one single injection per fraction. The details of the procedure are described in a previous publication [20].

In September 2011, the method was slightly extended: an additional fraction containing T-Ac and EpiT-Ac was collected during the HPLC-FC. This fraction needed a second HPLC clean up due to interference of T-Ac. During the first HPLC-FC, this T-Ac/EpiT-Ac fraction was always collected, but in order to keep the sample preparation as simple as possible, it only underwent the second HPLC-FC, GC-MS and GC-C-IRMS analysis if suspicious CIRs were determined for the primary compounds (A-Ac, Et-Ac,  $\beta\alpha\beta$ -Ac<sub>2</sub>,  $\alpha\alpha\beta$ -Ac<sub>2</sub> and PD-Ac<sub>2</sub>). In addition, the diethyl ether LLE was replaced by an *n*-pentane LLE. More details are available in a previous work [21]. At this stage, a new QC Neg was collected and stored at -80 °C. The new QC Neg comprised a volume of more than 3 L which allows us to use the same QC Neg for several years. If long-term shifts in CIR would occur, they can be detected and visualized with this QC Neg. In December 2011, a new QC Pos was made and stored at -80 °C for the same purpose. The new QC Pos was obtained by spiking QC Neg with sufficient amounts of A, Et,  $\beta\alpha\beta$ -diol,  $\alpha\alpha\beta$ -diol, T and EpiT.

**Table 1** Summary of the method evolution

February 2011	Launch routine method (A, Et, $\beta\alpha\beta$ -diol, $\alpha\alpha\beta$ -diol and PD)
September 2011	T, EpiT, <i>n</i> -pentane LLE, new QC Neg
December 2011	New QC Pos
February 2012	IRMS sensitivity $\times$ 10 (10 $\mu\text{L}$ solvent vent injections)
September 2012	$\delta^{13}\text{C}$ values acetate STD
May 2013	11-oxoEt+new calibration procedure
January 2014	QC STD MIX

In February 2012, we switched from 1- $\mu$ L splitless injections to 10- $\mu$ L solvent vent injections by using a programmed temperature vaporizer (PTV) on our GC-MS and GC-C-IRMS instrument. These 10  $\mu$ L solvent vent injections increased our IRMS sensitivity by a factor of 10 and allowed us to simplify our sample preparation procedure by reducing the required urine volume for IRMS analysis by a factor of 5. For a TC with a urinary concentration of 50 ng/mL, we now only needed 5 mL of urine instead of 25 mL in the past. More information is available in our previous work [21]. From now on, we consume 7 mL of QC Neg and 3 mL of QC positive per batch, prolonging the lifetime of our QCs.

From September 2012 on, we have started to actively record all  $\delta^{13}\text{C}$  values of all compounds in our acetate standard (acetate STD). The acetate STD is a mixture of acetylated standards of all our TCs, PD-Ac<sub>2</sub> and the IS (5 $\alpha$ -ol-Ac) and is analysed on GC-C-IRMS before, during and after every batch. It has always been included in our IRMS confirmation procedure, but was merely used to check and set the retention times of the compounds, estimate the GC chromatographic performance (peak tailing and separation) and evaluate the  $\delta^{13}\text{C}$  value of our IS as a deviating  $\delta^{13}\text{C}$  value could indicate problems during the combustion process.

Since May 2013, a second ERC (11-oxoEt) has been included in the method. The sample preparation procedure remained the same, but an additional fraction containing 11-oxoEt-Ac was collected during the HPLC-FC (6.6–8.1 min). This additional fraction was always collected, but in order to keep the sample preparation as simple as possible, it was only analysed on GC-MS and GC-C-IRMS if suspicious CIR were determined for the primary compounds or in another exceptional case. The IRMS calibration procedure was also adjusted. Instead of calibrating the CO<sub>2</sub> reference gas with an *n*-alkane mixture (C<sub>17</sub>–C<sub>25</sub>), CU-USADA 33-1 was used. During the calibration procedure, an additional check was performed using CU-USADA 34-1 and 5 $\alpha$ -androstane (IU 5 $\alpha$ -A), provided by the Biogeochemical Laboratories of the Indiana University (Bloomington, IN, USA). These adaptations were made to comply with the requirements of the new WADA technical document that would come into effect in September 2014 [17]. From January until September 2014, the guidelines document was in force, but both documents are essentially the same [22].

Since January 2014, an extra QC (QC STD MIX) was processed and analysed with every batch. QC STD MIX was prepared by spiking 3 mL of steroid stripped urine with 150  $\mu$ L of a 12.5  $\mu$ g/mL standard solution containing A, Et,  $\beta\alpha\beta$ -diol,  $\alpha\alpha\beta$ -diol, T, EpiT, PD and 11-oxoEt. QC STD MIX was added to the method in order to comply with WADA regulations [17]. However, a recent communication from WADA (dated 8th September 2014) has indicated that it is not necessary to subject reference materials with known CIR to the full sample preparation procedure, only that they must be injected prior to sample analysis.

## Routine doping samples

From February 2011 until May 2014, we analysed 609 routine urine samples from athletes on GC-C-IRMS. 279 samples (45.8 %) were investigated on IRMS because of an elevated T/E ratio, the other 330 samples (54.2 %) required IRMS investigation due to an elevated concentration of A, Et,  $\beta\alpha\beta$ -diol,  $\alpha\alpha\beta$ -diol, T or DHEA (or combinations) or because this was requested by the corresponding sport federation.

## Negative reference population samples

The collection and analysis of negative reference population samples occurred at different points in time during the period 2010–2013. In 2010, a first reference population consisting of 27 male (aged 26.0 $\pm$ 5.5 year) and 25 female (aged 26.3 $\pm$ 8.0 year) volunteers who declared not to have used any prohibited substance was analysed. The remainder of the negative reference population samples originated from healthy university staff that needed to undergo the yearly health check at the university hospital. These were collected and analysed in 2011–2013.

## Results and discussion

### Calibration

Ideally, the CO<sub>2</sub> reference pulse is calibrated against an isotopic standard with a chemical structure that is very similar or equal to that of the analytical compounds that need to be analysed on IRMS. It is important that reference materials have similar chemical properties because the combustion efficiency may vary, changing the isotopic composition of the gases [23]. To harmonize reported values and achieve uniform results between the different anti-doping laboratories worldwide, the IRMS system needs to be calibrated against a steroid reference material according to the new WADA regulations [17].

In 2008, two steroid isotopic standards (SIS), CU-USADA 33-1 and CU-USADA 34-1, have been assigned certified CIRs that are traceable to the international reference material [24]. CU-USADA 33-1 contains 5 $\alpha$ -ol-Ac, A-Ac, 11-oxoEt-Ac and 5 $\alpha$ -cholestane (5 $\alpha$ -Chol), CU-USADA 34-1 contains Et, A and PD (Table 2).

Since May 2013, our CO<sub>2</sub> reference pulse has been calibrated against CU-USADA 33-1 eight times. This SIS was chosen because it contains acetylated steroids, similar or identical to the ones we analyse in our GC-C-IRMS method. We calibrate with CU-USADA 33-1 in threefold and use the average assigned value as the new  $\delta^{13}\text{C}$  value for our CO<sub>2</sub> reference pulse. Immediately before and after the calibration,

**Table 2** Measured and certified CIRs for CU-USADA 33-1, CU-USADA 34-1 and IU 5 $\alpha$ -A

	$\delta^{13}\text{C}$ CU-USADA 33-1 (‰)				$\delta^{13}\text{C}$ CU-USADA 34-1 (‰)		
	5a-ol-Ac	5a-Chol	A-Ac	11-oxoEt-Ac	E	A	PD
	Before calibration ( $n=8$ )				$n=8$		
Average	-30.46	-24.80	-32.88	-17.92	-29.46	-27.26	-31.83
SD	0.17	0.28	0.39	0.66	0.43	0.57	0.62
Bias with certified value	<i>0.15</i>	<i>-0.03</i>	<i>0.16</i>	<i>-1.22</i>	<i>-0.55</i>	<i>-0.20</i>	<i>-0.34</i>
	After calibration ( $n=8$ )				Certified CIR		
Average	-30.69	-25.05	-33.31	-18.42	-28.91	-27.06	-31.49
SD	0.15	0.27	0.64	1.55	0.09	0.19	0.06
bias with certified value	<i>-0.08</i>	<i>-0.28</i>	<i>-0.27</i>	<i>-1.72</i>			
	total ( $n=16$ )				$\delta^{13}\text{C}$ IU 5 $\alpha$ -A (‰) ( $n=8$ )		
Average	-30.57	-24.92	-33.09	-18.17	-31.47		
SD	0.20	0.29	0.56	1.18	0.21		
bias with certified value	<i>0.04</i>	<i>-0.15</i>	<i>-0.05</i>	<i>-1.47</i>	<i>0.17</i>		
	Certified CIR				Certified CIR		
Average	-30.61	-24.77	-33.04	-16.7	-31.64		
SD	0.14	0.13	0.02	0.06	0.01		

Comments and conclusions presented in the manuscript are based upon these italic values

CU-USADA 33-1 is analysed as well to check how much the measured  $\delta^{13}\text{C}$  values of the steroids deviate from the certified  $\delta^{13}\text{C}$  values by using respectively the old and the new  $\delta^{13}\text{C}$  value of the  $\text{CO}_2$  reference pulse (Table 2). Afterwards, as an additional check, CU-USADA 34-1 and IU 5 $\alpha$ -A (5 $\alpha$ -androstane) are analysed with the new  $\text{CO}_2$  reference pulse  $\delta^{13}\text{C}$  value for the same purpose.

For CU-USADA 33-1, the bias with the certified values is acceptable, except for 11-oxoEt-Ac. For unknown reasons, the standard deviation (SD) and bias are unacceptable large for this compound. Similar observations were made for 11-oxoEt-Ac in CU-USADA 33-1 in a paper of Piper et al. [25]. The effect was partially attributed to an incomplete transfer from the GC column onto the hot zone of the combustion furnace. Other effects caused by the GC such as losses linked to the injection were also indicated as possible contributors to the atypical behaviour. An amount-dependent shift was also observed, but this does not apply for our measurements as we always inject the same amount during the calibration. 11-oxoEt-Ac is also present in our QC Neg, QC Pos and QC STD MIX but here we do not see a similar phenomenon and normal deviations are obtained (Table 3). The four steroids present in CU-USADA 33-1, including 11-oxoEt-Ac are used to set the  $\text{CO}_2$  reference pulse  $\delta^{13}\text{C}$  value, but despite this atypical behaviour of 11-oxoEt-Ac, a correct  $\delta^{13}\text{C}$  value for the reference pulse is obtained (acceptable bias for the other compounds). We assume that during the calibration process, the other three steroids compensate for the atypical behaviour of 11-oxoEt-Ac.

CU-USADA 34-1 consists of three non-acetylated standards. Free steroids cause more tailing during the GC analysis and this results in larger SDs. Overall, the bias is slightly worse in comparison with CU-USADA 33-1. The bias and SD for IU 5 $\alpha$ -A were in line with CU-USADA 33-1.

Calibrating with the new reference material CU-USADA 33-1 did not induce a shift in the  $\text{CO}_2$  reference gas value. The obtained  $\delta^{13}\text{C}$  values were in agreement with previously obtained values (using the alkane reference material). Neither was there a shift in the CIR of the IS, QC Neg and QC Pos (Figs. 1 and 2).

#### System stability

As recommended by a paper of Piper et al., the CIR of the IS has been monitored on a continuous basis as this compound is highly sensitive towards long term shifts and changes in the combustion process [26]. The data is presented in Fig. 1 with consecutive numbers instead of the date of measurement as data collection was carried out on a regular basis ( $n=579$ , three to four measurements per week). The average obtained value was  $-32.30\text{‰}$  with a SD of  $0.39\text{‰}$  and according to the Shapiro Wilk test ( $W=0.996$ ,  $p=0.221$ ) spread according to a normal distribution. The first 30 measurements were characterised by an average of  $-32.38\text{‰}$  and a SD of  $0.47\text{‰}$ , the last 30 measurements had an average of  $-32.23\text{‰}$  and SD of  $0.51\text{‰}$ . According to the independent  $t$  test at  $p=0.05$ , there is no significant difference between the first and last 30 measurements ( $p=0.236$ ), illustrating that no clear

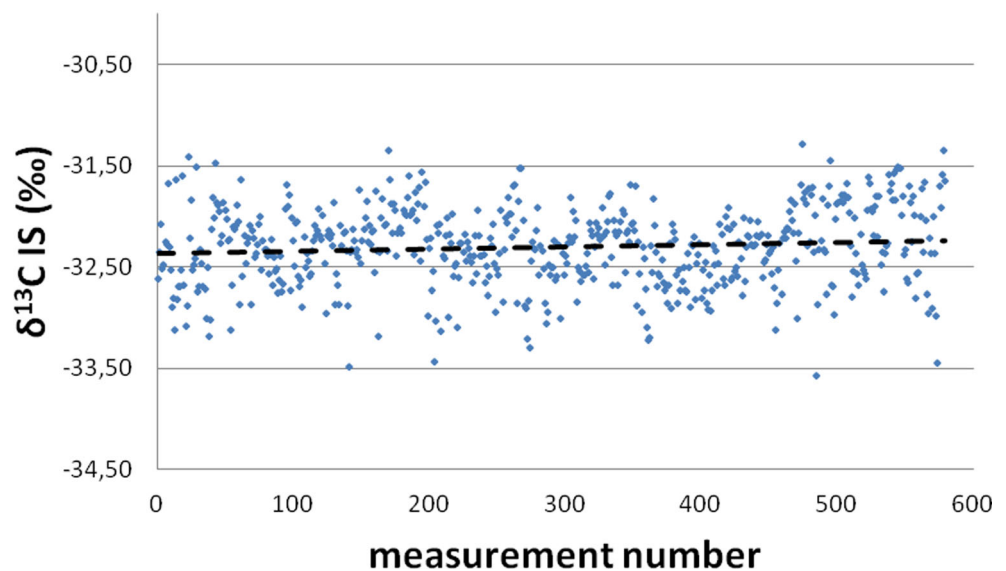
**Table 3** Statistical data gathered for QC Neg, QC Pos, QC STD MIX and the Acetate STD

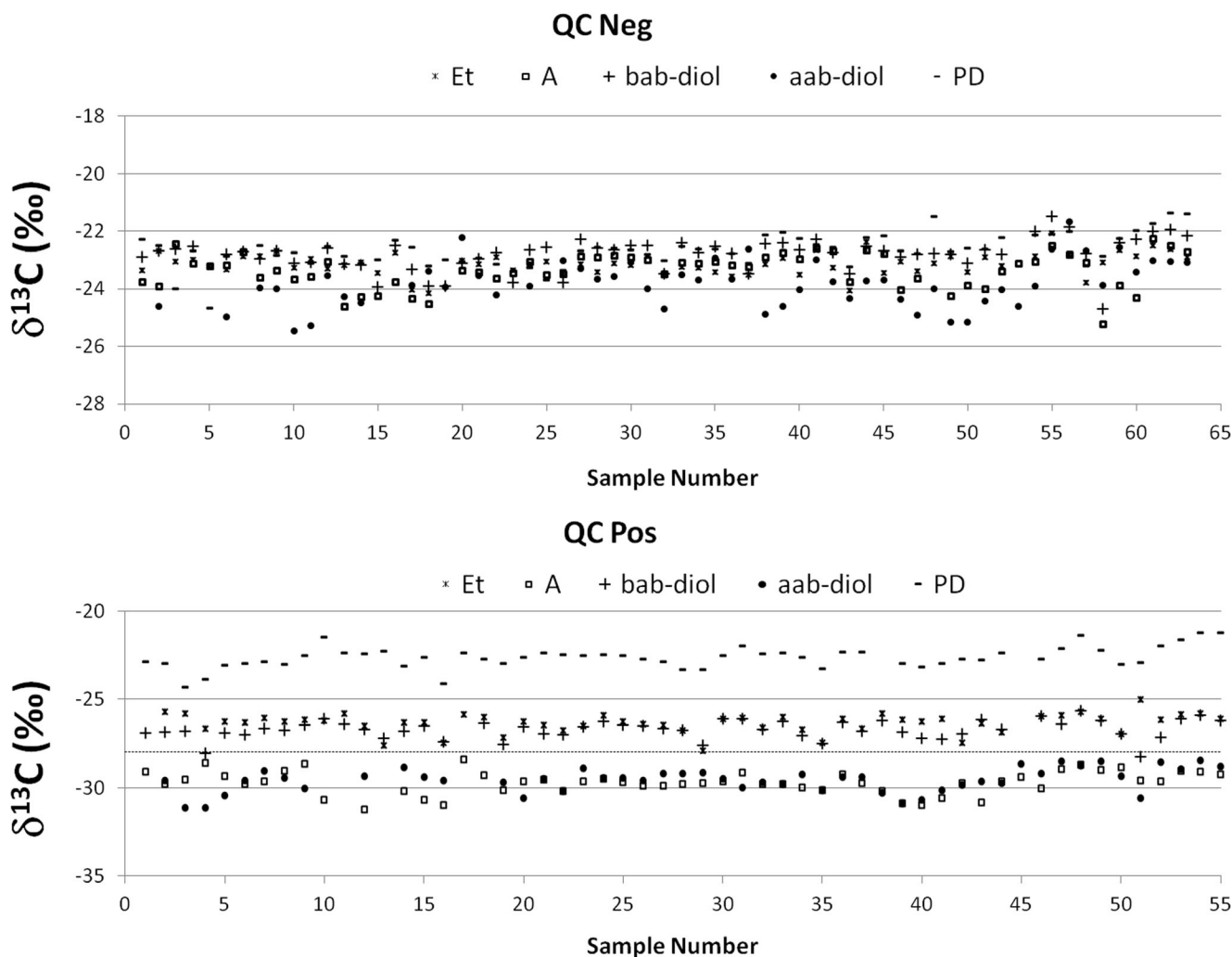
	$\delta^{13}\text{C}$ (‰)							
	Et	A	$\beta\alpha\beta$ -diol	$\alpha\alpha\beta$ -diol	EpiT	T	11-oxoEt	PD
QC Neg								
<i>n</i>	61	61	61	54	11	11	7	61
<i>u</i>	-23.1	-23.3	-22.8	-23.8	-25.3	-24.5	-22.9	-22.7
SD	0.38	0.61	0.56	0.81	0.78	0.69	0.31	0.56
<i>u</i> +2×SD	-22.4	-22.1	-21.7	-22.2	-23.7	-23.2	-22.3	-21.5
<i>u</i> -2×SD	-23.9	-24.6	-23.9	-25.4	-26.9	-25.9	-23.5	-23.8
QC Pos								
<i>n</i>	53	53	53	49	10	10	7	53
<i>u</i>	-26.4	-29.7	-26.7	-29.6	-28.9	-27.9	-23.0	-22.6
SD	0.54	0.65	0.53	0.68	0.87	0.83	0.45	0.62
<i>u</i> +2×SD	-25.3	-28.4	-25.7	-28.2	-27.1	-26.2	-22.1	-21.4
<i>u</i> -2×SD	-27.4	-31.0	-27.8	-30.9	-30.6	-29.5	-23.9	-23.9
QC STD MIX								
<i>n</i>	8	8	8	8	5	5	5	8
<i>u</i>	-30.0	-34.8	-33.7	-35.9	-34.9	-30.7	-19.7	-33.7
SD	0.49	0.47	0.19	0.64	0.42	0.19	0.36	0.63
<i>u</i> +2×SD	-29.0	-33.9	-33.3	-34.7	-34.0	-30.4	-18.9	-32.4
<i>u</i> -2×SD	-31.0	-35.8	-34.1	-37.2	-35.7	-31.1	-20.4	-34.9
Acetate STD								
<i>n</i>	137	137	137	137	137	137		137
<i>u</i>	-20.8	-35.8	-30.8	-36.5	-33.4	-32.8		-21.9
SD	0.61	0.42	0.57	0.61	0.49	0.61		0.35
<i>u</i> +2×SD	-19.6	-34.9	-29.7	-35.3	-32.4	-31.6		-21.2
<i>u</i> -2×SD	-22.0	-36.6	-32.0	-37.8	-34.3	-34.0		-22.7

significant shift occurred. Contrary to the expectations, the changes made to the method (different injection type and different calibration procedure) did not induce a clear shift.

In the publication of Piper et al., two factors contributed to two shifts: replacement of the CO<sub>2</sub> tank for the reference pulse and a change in the oxidation protocol for reoxidizing the

**Fig. 1** CIR of the IS from February 2011 until May 2014. The corresponding *dashed line* represents the Gaussian least-square fitted linear slope ( $y=0.0002x-32.36$ )





**Fig. 2** Quality control chart with CIR of Et, A, bab-diol, aab-diol and PD from QC Neg and QC Pos

combustion reactor. Since the launch of our method, it has not yet been necessary to replace the CO<sub>2</sub> tank. Our oxidization protocol has remained the same as well: we oxidize for 12 s before each measurement. If the IRMS has not been used for a longer period of time (>3 days), we start the batch with five blanks to ensure that the combustion reactor is saturated with enough oxygen before the actual measurements begin. We do not perform any other oxidation protocols (i.e. oxidize for a longer time period once a week). So far, we only had to replace the combustion reactor once (November 2011), but this did not induce a measurable shift either.

#### Quality control

##### *System blank*

So far, there has not been a case where the system blank indicated any problems or contaminations.

##### *QC Neg and QC Pos*

QC Neg and QC Pos were prepared and analysed with every batch of samples. The obtained CIR for the TCs and ERCs were plotted in a quality control chart and compared with the data acquired in the previous batches. If the CIR of one of the TCs or ERCs was outside the 2-sigma limits, all data from that batch was rejected and the batch was reanalysed [27]. The cause of the out-of-control data would be sought and solved and the batch would be reanalysed. Five times we have had out-of-control data: on one occasion, we were dealing with a broken combustion reactor and in a second incident we used a HPLC column that had exceeded its lifetime. The peaks had become so wide that they started to elute outside the collected fractions during HPLC-FC. This was especially the case for PD-Ac<sub>2</sub> (late eluting compound) and was more pronounced in QC Neg than QC Pos (same concentration of PD but a 7 mL aliquot compared to 3 mL for QC Pos). PD-Ac<sub>2</sub> had values of 0.5 and 0.3‰ outside the 2-sigma limit for QC Neg and QC

Pos respectively. On three other occasions, the GC column had exceeded its lifetime.

The collected data for QC Neg (starting from September 2011) and QC Pos (starting from December 2011) are presented in Table 3 and visualized in Fig. 2. For QC Neg, the largest SD was found to be 0.81‰ ( $\alpha\alpha\beta$ -diol), for QC Pos the largest SD was 0.87‰ (EpiT). Overall, the obtained precisions were in line with what can be expected for GC-C-IRMS measurements [26, 28]. Differences in precision between compounds are obviously caused by chemical and physical dissimilarities, which may for example lead to different chromatographic behaviour (i.e., tailing compounds are prone to more variation). However, differences in precision are also sample dependent. QC Pos is the same urine as QC Neg (i.e. same matrix) but spiked with the TCs. In QC Neg,  $\alpha\alpha\beta$ -diol is present in quite low concentrations, leading to a low peak intensity (but still within the acceptable linearity range) which will be susceptible to more variation. In QC Pos,  $\alpha\alpha\beta$ -diol is present in higher concentrations and has a lower SD (0.68‰). Likewise, a sample (and its TCs and ERCs) defined by a higher background will be prone to more variation. This is something one has to keep in mind when defining “the overall precision of a method” (cf. in the section “**Uncertainty**”).

#### QC STD MIX

QC STD MIX has basically the same function as QC Neg and Pos: It allows the analyst to identify deviating CIR for one or multiple compounds and to visualize long term shifts. QC STD MIX however has the advantage that the standards used to spike the steroid free urine are disposable and can also be analysed directly on GC-C-IRMS without sample preparation. The standards can be injected with or without acetylation which allows the analyst to for example identify acetylation problems if these would occur. QC STD MIX data is given in Table 3.

#### Acetate STD

The acetate STD has always been included in our IRMS confirmation procedure, but was merely used to check and set the retention times of the compounds, estimate the GC chromatographic performance (peak tailing and separation) and evaluate the  $\delta^{13}\text{C}$  value of our IS as a deviating  $\delta^{13}\text{C}$  value could indicate problems during the combustion process (cf. in the section “**System stability**”). When our IRMS method was launched, we assumed that if the  $\delta^{13}\text{C}$  value of the IS was within specifications, the GC-C-IRMS was working properly for  $\delta^{13}\text{C}$  value determination of the other compounds as well (if there were no chromatographic issues, impurities,...). However, in some occasions, the  $\delta^{13}\text{C}$  value of one or more acetylated standards started to deviate from their normal value and the deviation progressed in time whereas the  $\delta^{13}\text{C}$  values of the IS and the other acetylated standards remained correct,

indicating that incomplete combustion was not the problem. All compounds were still baseline separated, there was no peak tailing or other chromatographic issue and there were no visible impurities present. In these cases, replacement of the GC column however solved the issue and restored the  $\delta^{13}\text{C}$  values back to their normal values. We suspect that as the GC column gets older, (undetectable) column bleeding that elutes at specific retention times starts to influence the combustion process of compounds eluting at that retention time with a deviating  $\delta^{13}\text{C}$  value as a result. In our experience, the GC column lifetime rarely exceeds 600–700 injections for GC-C-IRMS purposes. However, to minimize costs, we transfer this “old” column to our GC-MS instrument as it still fulfils the chromatographic requirements for this analysis.

In this light, measuring the  $\delta^{13}\text{C}$  values of all compounds in our acetate STD facilitates troubleshooting to a great extent. Correct  $\delta^{13}\text{C}$  values for all compounds in our acetate STD guarantees that the GC-C-IRMS is working properly for CIR determination of all TCs and ERCs (and not only the IS). In the past, when deviating CIRs were obtained in the QC Neg and QC Pos, we were uncertain if the problem was situated in the sample preparation or originated from a malfunctioning GC-C-IRMS instrument. In our experience, in 90 % of the cases, deviating CIR are invoked by an “old” GC column. For the remaining 10 % there were various causes. One time they were caused by an HPLC column that had exceeded its lifetime. On a second occasion, deviating  $\delta^{13}\text{C}$  values were the result of a broken combustion reactor. In a third incident, deviating CIR resulted from a malfunctioning IRMS detector and an electronic control board needed to be replaced.

Some laboratories include additional quality controls in order to for example evaluate the efficiency of the hydrolysis or the derivatization. Because issues during the hydrolysis or derivatization will be traceable with our current quality controls, we have chosen not to include them in our already extensive method and we consider the presented strategy as sufficient.

#### Uncertainty

Since 2014, WADA demands a maximum combined measurement uncertainty ( $u_c$ ) of 1‰ [17]. To the best of our knowledge, only one publication has dealt with detailed uncertainty calculations for IRMS purposes [29]. These calculations however focused at certifying a reference material, rather than the combined measurement uncertainty of an IRMS method. The basic rule for the calculation of the combined uncertainty is the “square root of the sum of the squares rule”. Uncertainty components  $u(x_1) \dots u(x_n)$ , are combined as shown in the following equation [30]:

$$u_c = \sqrt{u(x_1)^2 + u(x_2)^2 + \dots + u(x_n)^2}$$

For GC-C-IRMS, there are no clear rules on how these separate uncertainty components need to be defined exactly. The consequence is that doping control laboratories use different methodologies to calculate their  $u_c$ , leading to a situation where some doping control laboratories have no problem maintaining their  $u_c$  beneath 1‰, whereas other

laboratories have great difficulties. In those cases, the apparent difference in combined uncertainty is not necessarily related to the quality of the GC-C-IRMS method, but caused by a difference in calculation methodology. An example of such a methodology might be the following (Table 4):

$$u_c = \sqrt{u(\text{calibration bias})^2 + u(\text{reference value})^2 + u(\text{method precision})^2}$$

The calibration bias is the difference between the certified value for the SIS and the measured SIS value as calculated in Table 2 (−0.15‰; the largest bias, disregarding 11-oxoEt-Ac). The second uncertainty component is the SD on the true certified SIS value as given in Table 2 (0.14‰; again the largest value). The largest contribution to  $u_c$  is the method precision that can for example be defined as the SDs for QC Neg (Table 3). As mentioned in section “QC Neg and QC Pos”, these SDs are not only component dependent, but also sample (and matrix) dependent. Laboratories that do not use real urine but spiked steroid stripped urine (or even water) for the calculation of this uncertainty component will most likely have lower SDs and a lower combined measurement uncertainty. Applying this methodology would, in our case, result in a maximum  $u_c$  of 0.84‰.

In general, the  $u_c$  will contain contributions from:

- 1) Precision of measurements

- 2) Bias of experimental processes
- 3) The uncertainty of CIR in reference materials used to fix and normalize the  $\delta$ -scale
- 4) The algorithms applied to correct and normalize the data [23].

The issue is that this requires the separate uncertainty components to be independent. IRMS works with values that are normalized in relation to VPDB, meaning that the uncertainty in normalized results for samples ( $\delta_{\text{true (sample)}}$ ) will have contributions from the precision of the measurements of the reference standards and sample, and the uncertainty in the reference value. This makes determination of the uncertainty of  $\delta_{\text{true (sample)}}$  complicated because of correlation. Fortunately, the correlation can be circumvented by calculating the uncertainty directly as described below [23].

The equation for calculating  $\delta_{\text{true (sample)}}$  can be written as:

$$\delta_{\text{true (sample)}} = \delta_{\text{true (RS1)}} + \left[ (\delta_{\text{raw (sample)}} - \delta_{\text{raw (RS1)}}) \times \left( \frac{(\delta_{\text{true (RS1)}} - \delta_{\text{true (RS2)}})}{(\delta_{\text{raw (RS1)}} - \delta_{\text{raw (RS2)}})} \right) \right]$$

RS1 and RS2 are the reference standards that were used for normalization using their measured and true values. In our GC-C-IRMS method, we have four RSs in CU-USADA 33-1, RS1 represents the steroid with the highest CIR and RS2 the steroid with the lowest CIR (or vice versa) so that the entire

calibration range is covered.  $\delta_{\text{true}}$  is the true CIR in relation to VPDB,  $\delta_{\text{raw}}$  is the non-normalized value.

For this equation, it is not possible to use the simple rules for combining uncertainties and the spreadsheet-based calculation as described by Kragten is the most straightforward

**Table 4** Combined uncertainty calculation according to the square root of the sum of the squares rule

	$\delta^{13}\text{C}$ (‰)							
	Et	A	$\beta\alpha\beta$	$\alpha\alpha\beta$	EpiT	T	11-oxoEt	PD
$u$ (calibration bias)	−0.15	−0.15	−0.15	−0.15	−0.15	−0.15	−0.15	−0.15
$u$ (reference value)	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
$u$ (method precision)	0.38	0.61	0.56	0.81	0.78	0.69	0.31	0.56
$u_c$	<i>0.43</i>	<i>0.64</i>	<i>0.60</i>	<i>0.84</i>	<i>0.81</i>	<i>0.72</i>	<i>0.37</i>	<i>0.60</i>

Comments and conclusions presented in the manuscript are based upon these italic values



approach [31]. The Kragten spreadsheet that applies for our IRMS method is presented in Table 5 and a clear explanation on how to set up these spreadsheets is given in the work of Carter and Barwick [23].

In Kragten sheet A, RS1 and RS2 are, respectively, 11-oxoEt-Ac and A-Ac from CU-USADA 33-1. Their  $\delta_{\text{true}}$  and its uncertainty can be found in Table 2,  $\delta_{\text{raw}}$  and its uncertainty were obtained from the same measurements as in Table 2, but before normalization with the reference value. For “sample”,  $\alpha\alpha\beta$ -Ac<sub>2</sub> of the QC Neg measurements was chosen (Table 3) as this compound was characterized by the largest SD. The  $\delta_{\text{raw}}$  was extracted from the data and the uncertainty calculated. The Kragten sheet calculates  $\delta_{\text{true (sample)}}$  and its  $u_c$ . Keep in mind that  $\delta_{\text{true (sample)}}$  is the value for  $\alpha\alpha\beta$ -Ac<sub>2</sub> whereas the value given in Table 3 is the corrected value for  $\alpha\alpha\beta$ -diol. The original non-corrected value for  $\alpha\alpha\beta$ -Ac<sub>2</sub> was  $-28.88\%$ ,

meaning that there is a difference of 0.65‰ between the value calculated by the Kragten sheet and the value generated by the IRMS software. This is likely caused by the IRMS software that uses all four steroids in CU-USADA 33-1 and in addition we are uncertain on how the conversion algorithms functions exactly.

In Kragten sheet B,  $\alpha\alpha\beta$ -Ac<sub>2</sub> has been replaced by 11-oxoEt-Ac from QC Neg (lowest SD). The original non-corrected value for 11-oxoEt-Ac was  $-25.75\%$ , meaning that there is a difference of 0.79‰ between the value calculated by the Kragten sheet and the value generated by the IRMS software.

In Kragten sheet C and D, 11-oxoEt-Ac has been replaced by 5 $\alpha$ -Chol as RS1. As mentioned in section “Calibration”, 11-oxoEt-Ac showed atypical behaviour and replacement by 5 $\alpha$ -Chol should give a better representation of  $u_c$ . 5 $\alpha$ -Chol

**Table 5** Combined uncertainty calculation with a Kragten spreadsheet. A and B: RS1=11-oxoEt-Ac, RS2=A-Ac; C and D: RS1=5 $\alpha$ -Chol, RS2=A-Ac. A and C: sample= $\alpha\alpha\beta$ -Ac<sub>2</sub>; B and D: sample=11-oxoEt-Ac. The calculated  $u_c$  is given in italics

Parameter	Value ( $\delta^{13}\text{C}, \%$ )	Uncertainty ( $\delta^{13}\text{C}, \%$ )						
A								
$\delta_{\text{true (RS1)}}$	-16.70	0.06	-16.64	-16.70	-16.70	-16.70	-16.70	-16.70
$\delta_{\text{true (RS2)}}$	-33.04	0.02	-33.04	-33.02	-33.04	-33.04	-33.04	-33.04
$\delta_{\text{raw (RS1)}}$	13.62	1.11	13.62	13.62	14.72	13.62	13.62	13.62
$\delta_{\text{raw (RS2)}}$	-0.76	0.55	-0.76	-0.76	-0.76	-0.21	-0.76	-0.76
$\delta_{\text{raw (sample)}}$	3.47	0.68	3.47	3.47	3.47	3.47	3.47	4.16
$\delta_{\text{true (sample)}}$	-28.23	<i>0.96</i>	-28.21	-28.22	-28.57	-28.69	-27.46	-27.46
Difference			0.02	0.01	-0.34	-0.46	0.77	0.77
B								
$\delta_{\text{true (RS1)}}$	-16.70	0.06	-16.64	-16.70	-16.70	-16.70	-16.70	-16.70
$\delta_{\text{true (RS2)}}$	-33.04	0.02	-33.04	-33.02	-33.04	-33.04	-33.04	-33.04
$\delta_{\text{raw (RS1)}}$	13.62	1.11	13.62	13.62	14.72	13.62	13.62	13.62
$\delta_{\text{raw (RS2)}}$	-0.76	0.55	-0.76	-0.76	-0.76	-0.21	-0.76	-0.76
$\delta_{\text{raw (sample)}}$	6.35	0.23	6.35	6.35	6.35	6.35	6.35	6.58
$\delta_{\text{true (sample)}}$	-24.96	<i>0.72</i>	-24.93	-24.95	-25.54	-25.29	-24.69	-24.69
Difference			0.03	0.01	-0.58	-0.33	0.27	0.27
C								
$\delta_{\text{true (RS1)}}$	-24.77	0.13	-24.64	-24.77	-24.77	-24.77	-24.77	-24.77
$\delta_{\text{true (RS2)}}$	-33.04	0.02	-33.04	-33.02	-33.04	-33.04	-33.04	-33.04
$\delta_{\text{raw (RS1)}}$	7.12	0.38	7.12	7.12	7.49	7.12	7.12	7.12
$\delta_{\text{raw (RS2)}}$	-0.76	0.55	-0.76	-0.76	-0.76	-0.21	-0.76	-0.76
$\delta_{\text{raw (sample)}}$	3.47	0.68	3.47	3.47	3.47	3.47	3.47	4.16
$\delta_{\text{true (sample)}}$	-28.59	<i>0.80</i>	-28.52	-28.59	-28.80	-28.88	-27.88	-27.88
Difference			0.07	0.01	-0.20	-0.29	0.72	0.72
D								
$\delta_{\text{true (RS1)}}$	-24.77	0.13	-24.64	-24.77	-24.77	-24.77	-24.77	-24.77
$\delta_{\text{true (RS2)}}$	-33.04	0.02	-33.04	-33.02	-33.04	-33.04	-33.04	-33.04
$\delta_{\text{raw (RS1)}}$	7.12	0.38	7.12	7.12	7.49	7.12	7.12	7.12
$\delta_{\text{raw (RS2)}}$	-0.76	0.55	-0.76	-0.76	-0.76	-0.21	-0.76	-0.76
$\delta_{\text{raw (sample)}}$	6.35	0.23	6.35	6.35	6.35	6.35	6.35	6.58
$\delta_{\text{true (sample)}}$	-25.57	<i>0.44</i>	-25.46	-25.57	-25.92	-25.63	-25.33	-25.33
Difference			0.12	0.00	-0.34	-0.06	0.25	0.25

**Table 6** Overview of routine samples confirmed by GC-C-IRMS

	Number of samples	T/E>4	Number of female samples	Negative	Positive
2011	209	63 (30.1 %)	25 (12.0 %)	206	3 (1.4 %)
2012	171	74 (43.3 %)	20 (11.7 %)	170	1 (0.6 %)
2013	157	86 (54.8 %)	26 (16.6 %)	154	3 (1.9 %)
2014 (until May)	72	56 (77.8 %)	26 (36.1 %)	71	1 (1.4 %)
Total	609	279 (45.8 %)	97 (15.9 %)	601	8 (1.3 %)

was chosen as it has the second highest CIR of CU-USADA 33-1. In sheet C, “sample” is again  $\alpha\alpha\beta$ -Ac<sub>2</sub> and in sheet D 11-oxoEt-Ac. The difference between  $\delta_{\text{true (sample)}}$  calculated by the Kragten sheet and the value generated by the IRMS software now drops to 0.29 and 0.17‰ for  $\alpha\alpha\beta$ -Ac<sub>2</sub> and 11-oxoEt-Ac, respectively.

In all four sheets  $u_c$  remains beneath 1‰ (0.96, 0.72, 0.80 and 0.44‰ for A, B, C and D respectively). Sheet C however gives the most representative value.

#### Routine doping samples

Table 6 gives an overview of the number of samples that were analysed from February 2011 until May 2014. 97 out of 609 samples (15.9 %) were female. Eight out of 609 samples (1.3 %) were positive, all of these were males. From 2011 until 2013, 40 % of all IRMS investigations were due to an elevated T/E ratio (>4). In 2014 almost 80 % of all IRMS requests were triggered by a T/E ratio above 4. This increase was caused by the new WADA regulations that demanded that urine samples with a T/E above 4 in the screening would be analysed on IRMS (after approval of the corresponding sport federation). The old WADA rules required that an elevated T/E (>4) was first confirmed with a second independent analysis before the corresponding sport federation received a request whether IRMS needed to be performed on the suspicious sample or not [32]. Taking the measurement uncertainty of

the T/E confirmation analysis in account, this meant that only T/E ratios above 4.43 were labelled as suspicious.

The highest T/E that resulted in a negative IRMS result was 18.5. For this sample, all  $\Delta\delta^{13}\text{C}$  values were between -0.2 and 0.4‰ and all  $\delta^{13}\text{C}$  values were above -23.4‰, making it unlikely that this was a false negative as the most enriched synthetic preparation that has ever been reported had a CIR of -22.9‰ [4, 5, 33, 34]. All T/E>19 were due to anabolic steroid abuse (Table 7). One sample with a T/E beneath 4 (T/E=2.4) led to an adverse analytical finding. All other adverse analytical findings had T/E ratios between 10 and 83. From 2011 until 2013, 30 % (195 out of 609) of all IRMS requests were triggered by a T/E between 4 and 6; in 2014 this number increased to 50 % (37 out of 72), but in 4 years time none of these samples exhibited  $\Delta\delta^{13}\text{C}$  values above the WADA IRMS threshold. These data support the proposition of Mareck et al. to raise the T/E threshold to 6 and only recommend an IRMS analysis for samples showing T/E>6 [19]. This would reduce the time and money spent on negative IRMS confirmations considerably and increase the overall efficiency of the IRMS confirmation procedure.

Recently, WADA has implemented the steroidal module of the athlete biological passport (ABP) [35]. The steroidal module monitors selected urinary steroid concentrations of an athlete over time in order to detect steroid doping. With each measurement, the ABP progressively adapts the calculated upper and lower reference limits, evolving from a population-based to an individual-based threshold [36]. As

**Table 7** Steroid profile and CIR of the positive samples

	Urinary concentrations (ng/mL)					$\delta^{13}\text{C}$ (‰)						
	Et	A	$\beta\alpha\beta$	$\alpha\alpha\beta$	T/E	Et	A	$\beta\alpha\beta$	$\alpha\alpha\beta$	T	11-oxoEt	PD
1	1717	1504	100	48	<i>21.8</i>	-29.4	-28.6	-29.6	-30.7			-21.4
2	4774	6046	127	526	<i>10.2</i>	-29.3	-29.3	-24.2	-30.4			-21.6
3	2660	2518	884	295	<i>22.6</i>	-27.1	-29.9	-27.5	-30.6			-24.1
4	2055	3534	189	15	<i>12.6</i>	-24.9	-24.5	-25.0	-26.8	-29.8		-23.0
5	2455	3082	222	93	<i>30.0</i>	-27.8	-27.8	-27.8	-27.6	-29.5	-24.7	-23.7
6	7828	8929	1306	673	<i>82.2</i>	-29.0	-29.9	-28.8	-31.9	-29.7	-23.4	-25.0
7	<i>17168</i>	7576	1479	43	2.4	-31.3	-28.5	-31.5	-30.3			-23.1
8	1524	1285	316	55	<i>81.1</i>	-27.4	-26.9	-27.8	-27.6			-18.0

These italic values show that we are dealing with positive samples

**Table 8** Overview of CIR obtained for the negative routine samples

	$\delta^{13}\text{C}$ (‰)								
	Et	A	$\beta\alpha\beta$	$\alpha\alpha\beta$	11-oxoEt	PD	A-Et	$\alpha\alpha\beta$ - $\beta\alpha\beta$	11-oxoEt-PD
Count	601	601	595	564	37	601	601	561	37
average	-23.9	-23.1	-23.9	-24.0	-23.7	-22.9	0.8	-0.1	-0.8
SD	1.1	1.1	1.2	1.1	1.1	1.0	0.9	0.9	1.1
Max	-19.8	-19.0	-18.5	-18.3	-20.4	-18.0	3.2	3.2	0.9
Min	-27.1	-26.9	-27.0	-27.2	-26.1	-26.7	-2.9	-3.2	-3.7
	$\Delta\delta^{13}\text{C}$ (‰)								
	PD-Et	PD-A	PD- $\beta\alpha\beta$	PD- $\alpha\alpha\beta$	11-oxoEt-Et	11-oxoEt-A	11-oxoEt- $\beta\alpha\beta$	11-oxoEt- $\alpha\alpha\beta$	
Count	601	601	595	564	37	37	37	33	
average	1.0	0.2	1.0	1.1	0.2	-0.3	0.2	0.7	
SD	0.8	0.8	0.8	0.8	1.1	1.1	1.2	1.1	

the number of test records increases, the calculated reference limits adapt from population thresholds (i.e.  $T/E > 4$ ) towards individual thresholds resulting in a more sensitive criterion. Over time, the use of the ABP will lead to a change in the types of samples submitted to IRMS analysis. More  $T/E < 4$  samples will undergo an IRMS analysis, while other ones with  $T/E > 4$  will no longer require testing. This fundamental change will be a big factor for IRMS analysis moving forward.

Table 7 gives an overview of all positive samples. These adverse analytical findings were conclusive and according to WADA guidelines. There were other samples that were suspicious and had one or multiple values close to the 3‰ threshold. Nonetheless, these cases did not fulfil the criteria and were reported as negative.

As pointed out in our previous publication,  $\beta\alpha\beta$ -diol and  $\alpha\alpha\beta$ -diol are more sensitive parameters than A and Et [20]. This can be attributed to lower endogenous dilution, longer detection times and a higher biological sensitivity due to their position in the metabolic pathway. In all positive samples, at least one of the androstane diols had a higher  $\Delta\delta^{13}\text{C}$  value than Et and A; and in all samples at least one of the

androstane diols had a  $\Delta\delta^{13}\text{C}$  value above 3‰. In sample number 4, for example, neither A or Et has a  $\Delta\delta^{13}\text{C}$  value above 3‰ whereas  $\alpha\alpha\beta$ -diol does. This illustrates and confirms that measuring only A and Et and not the androstane diols during the IRMS confirmation deteriorates the efficiency of the analysis.

Table 8 gives an overview of the collected data for the negative urine samples. The data for T and EpiT have been omitted as these TCs were only measured in cases that resulted in suspicious IRMS results in the first place. They were eventually labelled as negative but there is a high chance for false negatives within these samples. For the remaining TCs and ERCs it is obviously impossible to exclude the occurrence of any false negatives.

In a reference population with 56 subjects (36 males and 20 females), Flenker et al. pointed out that the difference between PD and A had an approximately symmetrical scatter around zero. In contrast, Et was on average  $\pm 1\%$  more depleted in relation to PD and A [7]. The same phenomenon is observed in our data. The effect was attributed to isotopic fractionation between the  $5\alpha$ - and  $5\beta$ -pathway, where the  $5\beta$ -compound becomes more depleted. Kinetic isotope effects originate from

**Table 9** Overview of CIR obtained for the negative reference population samples

	$\delta^{13}\text{C}$ (‰)								
	Et	A	$\beta\alpha\beta$	$\alpha\alpha\beta$	11-oxoEt	PD	A-Et	$\alpha\alpha\beta$ - $\beta\alpha\beta$	11-oxoEt-PD
Count	121	120	102	87	64	121	120	85	64
Average	-24.1	-23.5	-24.5	-24.3	-23.4	-23.5	0.7	0.1	-0.6
SD	1.0	0.9	1.3	1.2	1.0	1.2	0.7	0.9	0.9
Max	-21.4	-21.3	-21.3	-21.6	-21.5	-20.8	2.9	1.9	1.3
Min	-27.1	-25.8	-27.7	-27.2	-25.7	-26.9	-1.8	-3.2	-2.8
	$\Delta\delta^{13}\text{C}$ (‰)								
	PD-Et	PD-A	PD- $\beta\alpha\beta$	PD- $\alpha\alpha\beta$	11-oxoEt-Et	11-oxoEt-A	11-oxoEt- $\beta\alpha\beta$	11-oxoEt- $\alpha\alpha\beta$	
Count	121	120	102	87	64	64	58	54	
Average	0.7	0.0	1.0	0.9	0.4	-0.1	0.6	0.6	
SD	0.8	1.0	0.7	0.8	0.9	0.9	1.0	1.1	

either rate limitation or branching (or a combination). Reduction of the A-ring is rate limiting and generates two branches: 5 $\alpha$ - and 5 $\beta$ -metabolites [37]. However, our data shows only marginal difference between the average CIR of  $\alpha\alpha\beta$ -diol and  $\beta\alpha\beta$ -diol (0.1%). The average CIR for Et,  $\alpha\alpha\beta$ -diol and  $\beta\alpha\beta$ -diol are very similar to each other. The average CIR for A and PD are very similar as well but  $\pm 1\%$  more enriched in relation to the androstanediols and Et. Piper et al. analysed a population twice (one time with an old and one time with an optimized new HPLC-FC) [25]. Interestingly, with the old HPLC-FC, a difference of  $\pm 1\%$  between  $\alpha\alpha\beta$ -diol and  $\beta\alpha\beta$ -diol was observed. With their optimized HPLC-FC, the average CIR for E,  $\alpha\alpha\beta$ -diol and  $\beta\alpha\beta$ -diol were very similar to each other as is the case in our population and A and PD were also  $\pm 1\%$  more enriched, illustrating that there can be analytical causes for observed dissimilarities between laboratories.

In a study of Cawley et al. with a reference population of 167 subjects, the average CIR of Et was 0.3% depleted in relation to A [6]. The average  $\Delta\delta^{13}\text{C}$  value for Et and A with 11-oxoEt was 2.3 and 2%, whereas our population resulted in 0.2 and  $-0.3\%$ , respectively. The data of Cawley suggest that 11-oxoEt is enriched with  $\pm 2\%$  in comparison with the TCs making it less convenient to use as ERC in combination with a WADA 3% threshold. In our population however 11-oxoEt is suitable as ERC as 11-oxoEt is on average 0.8% depleted in relation to PD.

At this stage, it is unclear on how the parallels and differences between laboratories and populations need to be clarified in a consistent manner. The origin could be biological and physiological and related to the population. There could also be an analytical cause, resulting in (small) deviations between laboratories. In this context, analysing an identical small population (10–20 samples) by different laboratories might bring clarification. In any case, it is clear that every doping laboratory needs to verify the reference population limits for its relevant TCs and ERCs with their own reference population. Table 9 gives an overview of the collected data for the negative reference population and these values (average and SD) are very close to the values gathered with the negative routine samples.

## Conclusions

Our GC-C-IRMS confirmation method was performed on 609 routine doping samples. The data was investigated, presented and interpreted, covering different areas relevant to doping control. Special emphasis was given to the calibration, quality control and measurement uncertainty, in light of the 2014 WADA technical document. The paper concentrates on how IRMS results remain similar by switching to the 2014 WADA

technical document and gives insight in a straightforward approach to calculate the measurement uncertainty.

The method revealed eight cases of endogenous anabolic steroid abuse. The gathered data proves that using a T/E ratio threshold of 6 instead of 4 would increase the efficiency of the IRMS confirmation procedure. The introduction of the steroidal module of the ABP however will most likely invoke a fundamental change in the types of samples that will require IRMS analysis in the future. In addition, we demonstrated the reliability and validity of the GC-C-IRMS confirmation protocol.

## References

- Craig H (1957) Isotopic standards for carbon and oxygen and correction factors for mass-spectrometric analysis of carbon dioxide. *Geochim Cosmochim Acta* 12:133–149
- Becchi M, Aguilera R, Farizon Y, Flament MM, Casabianca H, James P (1994) Gas chromatography/combustion/isotope-ratio mass spectrometry analysis of urinary steroids to detect misuse of testosterone in sport. *Rapid Commun Mass Spectrom* 8:304–308
- Shackleton CHL, Phillips A, Chang T, Li Y (1997) Confirming testosterone administration by isotope ratio mass spectrometric analysis of urinary androstanediols. *Steroids* 62:379–387
- Ueki M, Okano M (1999) Analysis of exogenous dehydroepiandrosterone excretion in urine by gas chromatography/combustion/isotope ratio mass spectrometry. *Rapid Commun Mass Spectrom* 13:2237–2243
- de la Torre X, Gonzalez JC, Pichini S, Pascual JA, Segura J (2001) C-13/C-12 isotope ratio MS analysis of testosterone, in chemicals and pharmaceutical preparations. *J Pharm Biomed Anal* 24:645–650
- Cawley AT, Kazlauskas R, Trout GJ, Rogerson JH, George AV (2005) Isotopic fractionation of endogenous anabolic androgenic steroids and its relationship to doping control in sports. *J Chromatogr Sci* 43:32–38
- Flenker U, Güntner U, Schänzer W (2008) Delta13C-values of endogenous urinary steroids. *Steroids* 73:408–416
- Cawley AT, Trout GJ, Kazlauskas R, Howe CJ, George AV (2009) Carbon isotope ratio (delta13C) values of urinary steroids for doping control in sport. *Steroids* 74:379–392
- Piper T, Opfermann G, Thevis M, Schänzer W (2010) Determination of (13) C/(12) C ratios of endogenous urinary steroids excreted as sulpho conjugates. *Rapid Commun Mass Spectrom* 24:3171–3181
- Piper T, Geyer H, Gougoulidis V, Flenker U, Schanzer W (2010) Determination of (13) C/(12) C ratios of urinary excreted boldenone and its main metabolite 5 beta-androst-1-en-17 beta-ol-3-one. *Drug Test Anal* 2:217–224
- Wang J, Wu M, Liu X, Xu Y (2011) Profiling of urinary steroids by gas chromatography-mass spectrometry detection and confirmation of androstenedione administration using isotope ratio mass spectrometry. *Steroids* 76:1560–1565
- de la Torre X, Colamonici C, Curcio D, Molaioni F, Pizzardi M, Botrè F (2011) A simplified procedure for GC/C/IRMS analysis of underivatized 19-norandrosterone in urine following HPLC purification. *Steroids* 76:471–477
- Piper T, Fusshöller G, Emery C, Schänzer W, Saugy M (2012) Investigations on carbon isotope ratios and concentrations of urinary formestane. *Drug Test Anal* 4:942–950

14. Piper T, Emery C, Saugy M (2011) Recent developments in the use of isotope ratio mass spectrometry in sports drug testing. *Anal Bioanal Chem* 401:433–447
15. Cawley AT, Flenker U (2008) The application of carbon isotope ratio mass spectrometry to doping control. *J Mass Spectrom* 43:854–864
16. Van Eenoo P, Van Gansbeke W, De Brabanter N, Deventer K, Delbeke FT (2011) A fast, comprehensive screening method for doping agents in urine by gas chromatography-triple quadrupole mass spectrometry. *J Chromatogr A* 1218:3306–3316
17. WADA (2014) TD2014IRMS Detection of synthetic forms of endogenous anabolic androgenic steroids by GC-C-IRMS
18. Donike M, Barwald K, Klostermann K, Schänzer W, Zimmerman J (1983) The detection of exogenous testosterone. *Int J Sports Med* 4: 68
19. Mareck U, Geyer H, Fuschöller G, Schwenke A, Haenelt N, Piper T, Thevis M, Schänzer W (2010) Reporting and managing elevated testosterone/epitestosterone ratios—novel aspects after five years' experience. *Drug Test Anal* 2:637–642
20. Van Renterghem P, Polet M, Brooker L, Van Gansbeke W, Van Eenoo P (2012) Development of a GC/C/IRMS method—confirmation of a novel steroid profiling approach in doping control. *Steroids* 77:1050–1060
21. Polet M, Van Gansbeke W, Deventer K, Van Eenoo P (2013) Development of a sensitive GC-C-IRMS method for the analysis of androgens. *Biomed Chromatogr* 27:259–266
22. WADA (2014) World anti-doping program guidelines detection of synthetic forms of endogenous anabolic androgenic steroids by GC-C-IRMS
23. Carter JF, Barwick VJ (2011) Good practice guide for isotope ratio mass spectrometry. FIRMS
24. Zhang Y, Tobias HJ, Brenna JT (2009) Steroid isotopic standards for gas chromatography-combustion isotope ratio mass spectrometry (GCC-IRMS). *Steroids* 74:369–378
25. Piper T, Emery C, Thomas A, Saugy M, Thevis M (2013) Combination of carbon isotope ratio with hydrogen isotope ratio determinations in sports drug testing. *Anal Bioanal Chem* 405: 5455–5466
26. Piper T, Flenker U, Mareck U, Schänzer W (2009)  $^{13}\text{C}/^{12}\text{C}$  ratios of endogenous urinary steroids investigated for doping control purposes. *Drug Test Anal* 1:65–72
27. Taylor JK (1987) In: quality assurance of chemical measurements, Lewis Publishers
28. Aguilera R, Chapman TE, Pereira H, Oliveira GC, Illanes RP, Fernandes TF, Azevedo DA, Neto FA (2009) Drug testing data from the 2007 Pan American Games:  $\delta^{13}\text{C}$  values of urinary androsterone, etiocholanolone and androstanediols determined by GC/C/IRMS. *J Steroid Biochem Mol Biol* 115:107–114
29. Munton E, Liu FH, Murby EJ, Hibbert DB (2012) Certification of steroid carbon isotope ratios in a freeze-dried human urine reference material. *Drug Test Anal* 4:928–933
30. JCGM (2008) Evaluation of measurement data—guide to the expression of uncertainty in measurement. JCGM
31. Kragten J (1994) Tutorial review. Calculating standard deviations and confidence intervals with a universally applicable spreadsheet technique. *Analyst* 119:2161
32. WADA (2011) Guidelines for the reporting and management of elevated T/E ratios and endogenous steroids
33. Brooker L, Cawley A, Drury J, Edey C, Hasick N, Goebel C (2014) Stable carbon isotope ratio profiling of illicit testosterone preparations—domestic and international seizures. *Drug Test Anal* 6:996–1001
34. Cawley A, Collins M, Kazlauskas R, Handelsman DJ, Heywood R, Longworth M, Arenas-Queralt A (2010) Stable isotope ratio profiling of testosterone preparations. *Drug Test Anal* 2:557–567
35. WADA (2014) TD2014EAAS Endogenous anabolic androgenic steroids measurement and reporting
36. Sottas PE, Baume N, Saudan C, Schweizer C, Kamber M, Saugy M (2007) Bayesian detection of abnormal values in longitudinal biomarkers with an application to T/E ratio. *Biostatistics* 8:285–296
37. Schänzer W (1996) Metabolism of anabolic androgenic steroids. *Clin Chem* 42:1001–1020