Evaluation of uncertainty sources in the determination of testosterone in urine by calibration-based and isotope dilution quantification using ultra high performance liquid chromatography tandem mass spectrometry

J. Pitarch-Motellón, J. V. Sancho, M. Ibáñez, A. F. Roig-Navarro*. Research Institute for Pesticides and Water, Universitat Jaume I, E-12071, Castelló, Spain. E-mail: roig@uji.es

N. Fabregat-Cabello. Department of Clinical Chemistry, University of Liège, Liège, Belgium

O. J. Pozo. Bioanalysis Research Group, IMIM, Hospital del Mar Medical Research Institute. Barcelona, Spain

Rosa Ventura, Barcelona Antidoping Laboratory, Doping Control Research Group, Fundació IMIM, Barcelona, Spain. rventura@imim.es

JI García Alonso, Pablo Rodríguez-González, Adriana González Gago, Amaia Ereño Artabe. Department of Physical and Analytical Chemistry, University of Oviedo, Oviedo, Spain.

Peter Van Eenoo, Koen Deventer. Ghent University, Department of Clinical Chemistry, Microbiology and Immunology, Doping Control Laboratory, Zwijnaarde, Belgium.

Yvette Dehnes, Sebastian Rzeppa. Norwegian Doping Control Laboratory, Oslo University Hospital, Oslo, Norway

ABSTRACT

Three quantification methodologies, namely calibration with internal standard (Cal-IS, non-weighted), weighted calibration with internal standard (wCal-IS) and isotope pattern deconvolution (IPD) have been used for the determination of testosterone in urine by LC-MS/MS. Uncertainty has been calculated and compared for the three methodologies through intra- and inter-laboratory reproducibility assays. IPD showed the best performance for the intra-laboratory reproducibility, with RSD and combined uncertainty values below 4% and 9% respectively. wCal-IS showed similar performance, while Cal-IS where not constant and clearly worse at the lowest concentration assayed (2 ng/mL) reaching RSD values up to 16%. The inter-laboratory assay indicated similar results although wCal-IS RSD (20%) was higher than IPD (10%) and Cal-IS get worse with RSD higher than 40% for the lowest concentration level. Uncertainty budgets calculated for the three procedures revealed that intercept and slope were the most important factors contributing to uncertainty for Cal-IS. The main factors for wCal-IS and IPD were the volumes of sample and/or standard measured.

INTRODUCTION

The use of drugs to enhance performance in sports is a well-known and documented issue. Despite the continuous introduction of new compounds, endogenous androgenic anabolic steroids (EAAS) are among the most popular doping agents[1–3]. EAAS determination still represents an important challenge due to the complexity to differentiate exogenous administration of endogenous substances. The goal requires collaborative efforts as well as advanced methodologies[1–7]. Longitudinal fluctuations measurement for a given athlete is nowadays regarded as the most effective approach to suspect the EAAS misuse. In this way, the steroidal profile of the Athlete Biological Passport (ABP) represents a powerful tool to reveal doping with endogenous compounds[1,3,6].

For most drugs, urine is the matrix generally used since it involves a non-invasive sampling procedure, large volumes are easily obtained, shows wide time windows and concentrations are high enough[1,6,7].

However, sample preparation is mandatory to ensure matrix effect attenuation and good sensitivity and selectivity. Usual treatment techniques such as solid phase extraction (SPE), liquid-liquid extraction (LLE) and simple matrix dilution are normally used. Due to its simplicity, efficiency and low cost, LLE at basic pH is still widely used in EAAS determination in urine samples[5,6]. Concerning identification and quantification, LC-MS based techniques –equipped with Electrospray Ionization source (ESI)- tend to replace GC-MS(/MS) – considered as the gold World Anti-doping Agency (WADA) standard for quantifications[8]- since the former shows suitable sensitivity and faster instrumental run time. Specially UHPLC-MS/MS with its demonstrated separation efficiency is considered the method of choice in doping analysis[1,5,6,9].

A relevant problem with the use of ESI source is the signal alteration due to matrix effect[10-12]. Matrix effect can affect drastically to sensitivity, precision and accuracy of the analytical results. The most robust approach to minimize matrix effect rely on the use of Stable Isotope Labeled-Internal Standard (SIL-IS)[11,12]. Thus, matrix-effects associated to complex matrices can be properly overcome using a quantification methodology based on isotope dilution mass spectrometry (IDMS). Classical IDMS is based on the preparation of methodological calibration curves with the associated time consumption. An alternative method of quantification, based on the measurement of isotopic abundances in the spiked sample by multiple linear regression, can also be used. This method, known as isotope pattern deconvolution (IPD), do not requires the construction of any calibration graph and has been tested satisfactorily for rapid quantifications in different complex matrices[13-16]. IDMS together with IPD is a fast and reliable methodology, which provides one result per injection with high accuracy and free of matrix effect.

In the field of doping analysis, improvements of reliability and robustness of analytical results is continuously and still required[1,2,5,6]. WADA highlights the need of good inter-laboratory precision, particularly relevant in ABP profiling[5]. Analytical results for ABP are obtained from different laboratories for the same athlete, thus, improving inter-laboratory precision seems of maximum concern to allow universal application of any developed methodology. In this way, the need of calculating and minimizing measurement uncertainty deserves to be treated thoroughly[2,17,18].

In the present work, a previously developed method has been applied to assess the uncertainty in the testosterone concentration determined in several synthetic urine samples. Testosterone concentration has been calculated using three different methodologies, weighted and non-weighted calibration with IS (wCal-IS and Cal-IS, respectively) and IPD. In order to evaluate more in depth the associated uncertainty, an inter-laboratory comparison among five laboratories has been performed. For all three methodologies, intra- and inter-laboratory measurements have been conducted, combined uncertainties (u_c) and full uncertainty budgets have been obtained and compared.

EXPERIMENTAL

Reagents and materials

Testosterone (T, purity 99%) was provided by Sigma-Aldrich Co. (Madrid, Spain) and $^{13}C_2$-testosterone ($^{13}C_2$-T, purity 98% and $^{13}C_2$-enrichment 98%) by Cambridge Isotope Laboratories (Andover, MA, USA).

Methanol (MeOH, HPLC quality) and methyl tert-butyl ether (MTBE, GC quality) were provided by Scharlau (Barcelona, Spain). For the sample hydrolysis, β-glucuronidase from *E. coli* K12 provided by Roche (Indianapolis, IN, USA) was employed. A 1 M phosphate buffer was prepared by dissolving the proper amount of $(NH_4)_2HPO_4$ (Merck, Darmstadt, Germany) in Milli-Q water and adjusted to pH=7 with HCl 37% from Scharlau (Barcelona, Spain). Also, a $NaHCO_3/Na_2CO_3$ (1:2, w/w) (Sigma-Aldrich Co., Madrid, Spain)

solid buffer was prepared. Formic acid (LC additive quality) and a 500 mM solution of NH_4HCOO (Scharlau, Barcelona, Spain) in methanol HPLC were used for the mobile phase preparation.

A 250 µg/mL stock solution of T was prepared by dissolving 25 mg of solid standard, accurately weighed, in 100 mL of methanol. The stock solution of $^{13}C_2$-T was prepared by dissolving 10 mg of the purchased material in 50 mL of methanol. This provided a concentration by reverse isotope dilution against the natural compound of 237 µg/mL.

Individual 10 µg/mL and 1 µg/mL working solutions of the natural and labelled compounds were prepared by dilution of the stock solutions with methanol. All of the standard solutions were stored in amber glass bottles in a freezer.

The water purification system used was a Milli-Q gradient A10 from Millipore (Bedford, MA, USA).

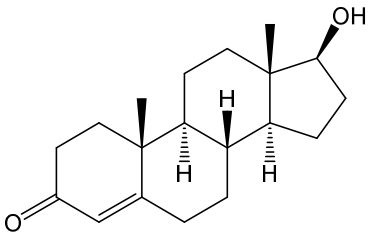
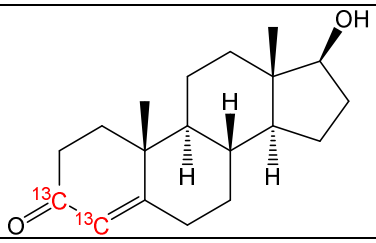
Instrumentation

All participants in the inter-laboratory comparison have determined testosterone by LC-MS/MS. Additionally some laboratories have used other methodologies (see inter-laboratory comparison section).

This section describes the instrumentation used at Research Institute for Pesticides and Water (IUPA) laboratory, where the intra-laboratory measurements and all calculations have been done.

An Acquity UPLC system coupled to a TQD triple quadrupole mass spectrometer from Waters Corp. (Milford, MA, USA) was employed for sample analysis. Chromatographic separation was performed with an Acquity UPLC BEH C18 column (1.7 µm, 2.1 mm x 100 mm), also from Waters Corp., at a 0.3 mL/min flow rate and an injection volume of 10 µL. The column oven was kept at 55ºC and the sample manager at 10ºC. Mobile phase A was purified water and mobile phase B was MeOH HPLC, both containing 0.01% of formic acid and 1 mM of NH_4HCOO as modifiers. The gradient applied was: 45% B (0-1 min), linear increase to 77.5% B in 6.5 min, 95% B (7.51-8 min), 45% B (8.5-11.5 min). Chromatograms of blank and a selected sample can be seen in Figure S.8 in supplementary material.

Table 1. Chemical structure and experimental conditions of the LC-(ESI)-MS/MS for testosterone and labeled testosterone

Compound	Structure	Rt (min)	Precursor ion	Cone voltage (V)	SRM transitions
T		5.7	$[M+H]^+$	30	**289.2 > 96.9** (25)
					290.2 > 96.9 (25)
					289.2 > 108.9 (25)
$^{13}C_2$-T		5.7	$[M+H]^+$	30	**291.2 > 98.9** (25)
					292.2 > 98.9 (25)

Electrospray ionization in the mass spectrometer was performed at 120 ºC and 350 ºC source and desolvation temperatures, 80 and 800 L/h cone gas and desolvation flow, respectively, and 3.5 kV capillary voltage, operating in positive ion mode. MS/MS experimental conditions for T and $^{13}C_2$-T are listed in Table 1.

Nitrogen was employed as both drying and nebulizing gas, obtained from a nitrogen generator N_2 LC-MS adapted for LC-MS analyzers (Claind, Teknokroma, Barcelona, Spain). Collision cell was operated under a pressure of approximately 5.6 x 10^{-3} mbar of argon 99.995% (Praxair, Madrid, Spain). Dwell times of 0.1 s per SRM transition were chosen. MassLynx v4.1 (Waters) and homemade Excel spreadsheets were used to process the data obtained. Relative abundances of individual 100 ng/mL standards were determined (n=5) under this conditions with RSD values under 1.5%.

Sampling and sample preparation

The aim of the study was explained to 15 healthy volunteers (8 men and 7 women with ages comprised between 16 and 59 years) and consent was obtained after confirmation that they fully understood the experiment. Urine samples were collected and stored at -20ºC until use. Testosterone concentration was approximately determined by IPD for all samples. 12 samples were selected and mixed in pairs in approximate 1:1 (v/v) ratios to obtain 6 synthetic urine samples, A to F, with increasing concentrations along the 2 ng/mL to 75 ng/mL testosterone range.

2.5 mL of the synthetic samples were transferred to individual glass tubes, together with 25 µL of 1 µg/mL $^{13}C_2$-T, and they were neutralized with 1 mL of 1 M phosphate buffer (pH 7.0). Then, 30 µL of β-glucuronidase solution were added. Samples were incubated at 55 ± 2 ºC in a water bath for 1 h.

After hydrolysis, approximately 200 mg of a $NaHCO_3/Na_2CO_3$ (1:2, w/w) solid buffer were added and dissolved by stirring in a vortex. Liquid-liquid extraction was performed with 6 mL of MTBE and stirring in a vortex for 1 min. Separation of phases was achieved by centrifugation at 3500 rpm for 5 min and the upper organic phase was transferred to clean glass tubes with disposable Pasteur pipettes, carefully avoiding to transfer any aqueous phase. MTBE was evaporated in a MiVac at 40ºC for 20 min. The residue was reconstituted in 300 µL of MeOH/H_2O 1:1 (v/v) and transferred to LC vials.

Samples and standards were equally treated and analyzed as described above.

Quantification methods

The 6 synthetic urine samples (A to F) were analyzed by three quantification approaches: Cal-IS, wCal-IS and IPD.

Additionally, at IUPA laboratory, standard addition was also employed for the inter-laboratory experiment. On this purpose, 2.5mL aliquots of each sample were spiked with 0, 0.5, 2 and 3.5 times the original approximate concentration of T and adjusted to a final volume of 2720 µL with water. The described sample treatment was applied without the addition of internal standard.

For all participant laboratories, calibration curves freshly prepared consisted in 6 points between 0 and 100 ng/mL of T in 2.5mL of water. Using the same data acquired for calibration with IS, weighed calibration calculations were applied as described in Garcia-Alonso and Rodríguez-González[19]. The weighing factor used has been the common value inverse of the variance ($1/SD^2$).

IPD was applied to the same sample extracts used in Cal-IS.

The isotope dilution quantification methodology employed is based on multiple linear regression and the spiking of samples with an isotopically enriched analog of the analytes of interest. This produces an intentional alteration of the natural isotopic composition of the analyte in the mix. Briefly, the altered isotopic composition measured in the mixture $A_{mix}^{SRM_i}$ is a combination of the contribution of the abundances of the natural, $A_{nat}^{SRM_i}$, and the isotopically enriched spike, $A_{lab}^{SRM_i}$ analyte. For a single isotopically enriched spike and *n* measured transitions, this can be expressed in matrix notation as:

$$\begin{bmatrix} A_{mix}^{SRM1} \\ A_{mix}^{SRM2} \\ \vdots \\ A_{mix}^{SRMn} \end{bmatrix} = \begin{bmatrix} A_{nat}^{SRM1} & A_{lab}^{SRM1} \\ A_{nat}^{SRM2} & A_{lab}^{SRM2} \\ \vdots & \vdots \\ A_{nat}^{SRMn} & A_{lab}^{SRMn} \end{bmatrix} \begin{bmatrix} X_{nat} \\ X_{lab} \end{bmatrix} + \begin{bmatrix} e^{SRM1} \\ e^{SRM2} \\ \vdots \\ e^{SRMn} \end{bmatrix}$$

A vector error e^{SRM1} needs to be included in order to solve the system by multiple linear regression, which gives the molar fractions of natural and labelled compounds (X_{nat} and X_{lab} respectively) as solutions. These can be obtained in any spreadsheet software with a linear regression function (LINEST in Microsoft Excel) inserting the data in matrix form. Then, since the added amount of labelled compound N_{lab} is known, the amount of natural compound in the sample N_{nat} is readily calculated (Table S.4. Supplementary information):

$$N_{nat} = N_{lab} \frac{X_{nat}}{X_{lab}}$$

As it can be seen, no methodological calibration procedures are required and a single injection provides one concentration value of the sample. The mass isotopomer distribution from each precursor ion has to be determined experimentally in the mix and compared with the individual distributions corresponding to the natural and the labeled analyte, the reference distributions. These individual distributions can be theoretically calculated or experimentally measured. In the present work, they have been experimentally measured. In a first step, theoretical isotopomer distributions have been obtained by IsoPatrn software implemented by L. Ramaley and L. Cubero-Herrera[20]. Afterwards, only those transitions producing instrumental signal significantly different from background have been selected. A thorough description of the general IPD methodology and its application to different analyte types can be found in the literature[21-23].

Inter-laboratory experiment

For the inter-laboratory variability evaluation of both calibration and IPD methods, four laboratories were contacted and agreed to collaborate: Barcelona Antidoping Laboratory (Fundació IMIM, Barcelona, Spain), Doping Control Laboratory (DoCoLab, Ghent University, Ghent, Belgium), Norwegian Doping Control Laboratory (Oslo University Hospital, Oslo, Norway) and the Department of Physical and Analytical Chemistry at University of Oviedo (Oviedo, Spain).

Three plastic bags containing 12 mL of frozen samples A to F in individual Falcon tubes, a vial with 1 mL of $^{13}C_2$-T 10 µg/mL in MeOH and Instructions and Results documents were prepared. The bags were put into sealed packages with the required amount of dry ice to ensure sub-zero temperature conditions until arrival to the selected laboratories. Samples were processed and all the required measurements were performed in order to apply calibration and IPD calculations at our laboratory. In addition, laboratories were also asked to perform any other routine quantification method they had implemented (Table 2). Taking into account those extra quantification methods, we got 19 analytical results for each sample. These

results were used to calculate a consensus value for the concentration of each sample, C_{ref}, and its associated uncertainty, u_{ref}.

Table 2. Additional quantification procedures conducted in inter-comparison participating laboratories

Laboratory	Additional analytical methods
IUPA	Standard additions (LC-MS/MS)
DoCoLab	GC-MS/MS, LC-HRMS
Norwegian Doping Control Laboratory	GC-MS/MS

Uncertainty assessment

In the present work, measurement of uncertainty was evaluated according to WADA technical document TD2014DL [24] and references therein. Specifically the procedure based in the Nordtest guide [25]. To this end, an intra-laboratory reproducibility experiment (five replicates of each sample along five weeks) and a short inter-laboratory comparison have been conducted. Combined uncertainty, u_{comb}, for each selected quantification method and sample were calculated and compared. Moreover, the inter-laboratory reproducibility standard deviation was calculated and compared for the three selected methods. Combined uncertainty has been calculated as:

$$u_{comb} = \sqrt{u_{SD}^2 + u_{bias}^2}$$

where u^2_{SD} is the intra-laboratory reproducibility standard deviation for the five replicates obtained along five consecutive weeks at IUPA laboratory and u_{bias} is the uncertainty associated to any source of bias which accounts for the method and laboratory bias, including the uncertainty associated to the consensus reference value. To that purpose, a short inter-laboratory comparison was conducted and a whole of 19 quantification results have been obtained for each sample A to F (see inter-laboratory experiment section). Thus, u_{bias} was calculated as

$$u_{bias} = \sqrt{RMS^2 + u_{ref}^2}$$

where u_{ref} is the bias uncertainty associated to the consensus concentration value for each sample, C_{ref}, obtained by:

$$u_{ref} = \frac{S_R}{\sqrt{n}}$$

where S_R is the mean standard deviation for the inter-laboratory reproducibility and *n* is the number of results for each sample. A *n*=17 was employed instead of 19 due to exclusion of outliers determined by Hampel test (see results, Table 3 and Table S.7 from Supplementary Information).

RMS is the root mean square bias for each quantification method used in the intra-laboratory reproducibility assessment conducted alt IUPA lab (for examples of calculations see Table S.5 in supplementary material).

For each sample (A to F) a mean bias has been calculated from the intra-lab reproducibility study (n = 5). These mean bias have been used to calculate RMS as:

$$RMS = \sqrt{\frac{\sum_i bias_i^2}{6}}$$

On the other hand, contribution of any source of uncertainty to a given measurement, known as full uncertainty budget, can be calculated using the Kragten approach [26]. Briefly, it consists in an approximation of error propagation theory calculations adapted for its implementation in spreadsheet programs (such as Microsoft's Excel). Calculation tables are constructed with all the parameters used to obtain the final analytical result including their uncertainty or standard deviation. Then, parameter values are sequentially altered with their SD to obtain the deviation (Δ^2) produced to the analytical result in relation to the unchanged value, which constitutes the magnitude of contribution to total uncertainty of the analytical procedure. It is readily calculated for each parameter *i* as:

$$\Delta_i^2 = (x - x_i)^2$$

Where *x* is the unchanged value and x_i is the new value with one parameter altered. Then, total uncertainty of the procedure (*U(x)*) can be obtained using:

$$U(x) = \sqrt{\sum_i \Delta_i^2}$$

Examples of complete uncertainty calculations can be consulted in the Supplementary Information (Table S.6).

RESULTS AND DISCUSSION

IPD measurements

As explained above, IPD calculations rely on the relative abundance distribution of natural and labelled compounds and, therefore, on their accuracy. For this purpose, the most abundant SRM transitions for each compound were selected with IsoPatrn software. Then, relative abundances were experimentally determined by preparing individual 100 µg/L standards in MeOH/H_2O 1:1 (v/v) and injecting them five times each (Tables S.1 and S.2 in supplementary information). Mean values for experimental abundances were used in subsequent quantification procedure and standard deviations were used in the uncertainty budgets building procedure.

IPD calculation also requires to know the exact amount of labelled compound added to samples. Exact concentration of the $^{13}C_2$-T working standard solution was calculated by reverse isotope dilution (RID) against the natural T solution, resulting in 12.20 ± 0.10 mg/L. (Table S.3 supplementary information).

Evaluation of uncertainty

Uncertainty has been assessed as intra-laboratory reproducibility standard deviation and through the reproducibility of an inter-laboratory comparison.

Intra-laboratory has been conducted at IUPA facilities. Five replicates of the synthetic urine samples (A to F) have been analyzed along five consecutive weeks. Concentration mean values, as well as standard deviation and RSD(%) are shown in Table 3.

Regarding intra-lab precision, wCal-IS shows RSD below 5% for any concentration level. IPD quantification performs slightly better while Cal-IS clearly achieves the worst reproducibility at the lower concentrations, reaching a value of 15.8 % at 2 ppb level. Concerning the inter-laboratory comparison, results are qualitatively similar. IPD shows the highest precision, with a mean RSD value around 10%, while wCal-IS doubles that value. On the other hand, Cal-IS shows the worst performance at the lowest levels, where RSD reaches values higher than 40%.

Table 3. Intra-laboratory and inter-laboratory precision data for the three quantification methodologies assayed and for the consensus value. Cal_IS: non-weighted calibration with internal standard; w-Cal_IS: weighted calibration with internal standard; IPD: isotope pattern deconvolution.

	Mean ± SD (RSD%) ng/mL						
Sample	**Intra-lab reproducibility (n=5)[a]**			**Inter-lab reproducibility (n=5)[b]**			**Consensus**
	Cal_IS	**w-Cal_IS**	**IPD**	**Cal_IS**	**w-Cal_IS**	**IPD**	**C_{ref} (n=17)[c]**
A	1.9 ± 0.3 (15.8)	2.04 ± 0.06 (2.9)	2.10 ± 0.04 (1.9)	2.1 ± 0.9 (43)	2.2 ± 0.4 (18)	2.3 ± 0.3 (13)	2.0 ± 0.4 (20)
B	3.6 ± 0.3 (8.3)	3.78 ± 0.14 (3.7)	3.87 ± 0.06 (1.6)	4.1 ± 1.0 (24)	4.2 ± 0.8 (19)	4.3 ± 0.4 (9.3)	4.0 ± 0.4 (10)
C	10.1 ± 0.4 (4.0)	10.3 ± 0.5 (4.9)	10.48 ± 0.22 (2.1)	11.2 ± 1.9 (17)	11.6 ± 2.3 (20)	11.6 ± 0.9 (7.8)	11.2 ± 1.2 (11)
D	17.1 ± 0.9 (5.3)	17.4 ± 0.8 (4.6)	17.7 ± 0.6 (3.4)	19.5 ± 2.3 (12)	20 ± 4 (20)	20.3 ± 2.2 (11)	20 ± 3 (15)
E	51.7 ± 1.3 (2.5)	52.0 ± 1.1 (2.1)	52.9 ± 0.9 (1.7)	57 ± 8 (14)	60 ± 13 (22)	60 ± 5 (8.3)	57 ± 6 (11)
F	67.8 ± 1.9 (2.8)	68.0 ± 1.6 (2.4)	69.2 ± 0.9 (1.3)	75 ± 9 (12)	77 ± 16 (21)	78 ± 7 (9.0)	76 ± 10 (13)

[a] Results from IUPA laboratory

[b] One result from each of the five participant laboratories

[c] Consensus value calculated as the mean value for the results obtained from all quantification methodologies used in the interlaboratory comparison. Outlier values were excluded following Hampel test (Table S.7). Associated bias uncertainty of C_{ref} (u_{ref}) calculated from the Mean RSD% is 3.1% (see Table S.5)

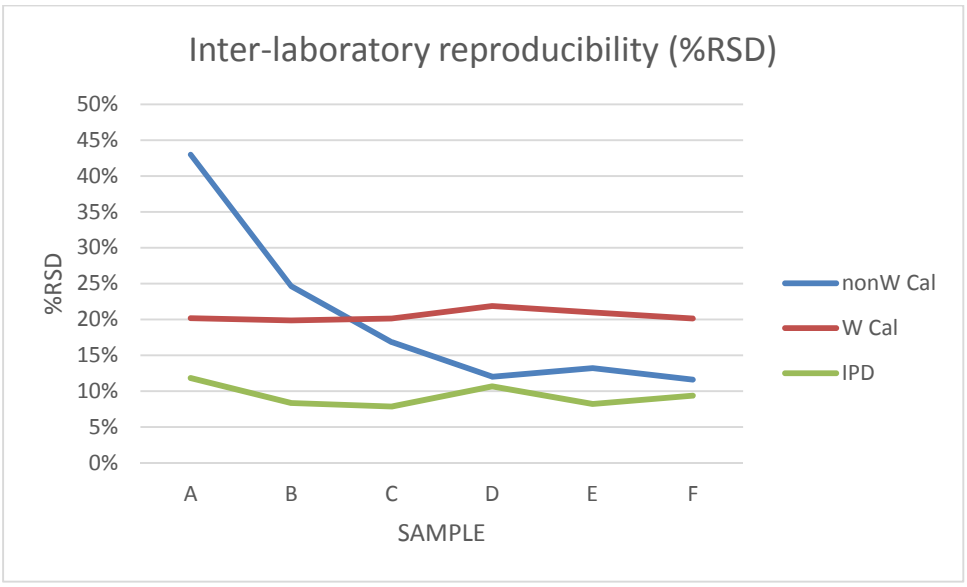


Figure 1. Mean inter-laboratory RSD values for the three quantification methods assayed. Concentration ranges from 2 ng/mL (sample A) to 70 ng/mL (sample F).

Thus, inter-laboratory reproducibility noticeably shows tendencies with concentration (Figure 1). Cal-IS performs poorly at low concentrations, with RSD > 40% in sample A, which decreases to values near 12% as concentrations get higher. In comparison, wCal-IS provided constant values of RSD along the concentration range (20%), improving uncertainty at low concentrations but performing slightly worse in the rest of the samples. IPD produced significantly lower dispersion of values resulting in the highest inter-laboratory precision (from 7.8% to 13%) of the three methods even at low concentrations.

In addition to intra-laboratory reproducibility, combined uncertainty, u_c, were calculated in order to estimate the measurement uncertainty for the three quantification methods. To this end, method and laboratory bias were estimated, according to the Nordtest guide [25] (see experimental), as the square root of two components: the percentage of the mean difference (RMS_{bias}) from a reference value (C_{ref}), and uncertainty of this reference value, u_{ref}. The end value for u_c accounts for the method and laboratory bias together with standard deviation of reproducibility at each concentration assayed (A-F samples) (Table 4).

Table 4. Combined uncertainty for the three quantification methodologies assayed.

Sample	Combined uncertainty (%)		
	Cal_IS	w-Cal_IS	IPD
u_{ref}		3.1%	
RMS_{bias}	10.4%	8.3%	6.9%
u_{bias}	10.9%	8.8%	7.6%
A	17.9%	9.4%	7.8%
B	13.2%	9.6%	7.8%
C	11.7%	10.0%	7.9%
D	12.0%	10.0%	8.4%
E	11.2%	9.1%	7.8%
F	11.3%	9.1%	7.7%

The consensus values obtained from the inter-laboratory comparison were adopted as reference values (Table 3). The consensus values are not intended to be used as certified values, but they were accepted as reference values to calculate bias uncertainty for each quantification methodology and to assess the bias associated to that reference value. A u_{ref} of 3.1% was obtained from the mean RSD value (12.9%) and *n*=17 from the 19 quantification procedures applied minus outlier values (see experimental section and Supplementary Information for details).

Since the data required for Cal-IS and wCal-IS is exactly the same, taking into account the difference in combined uncertainty (11.2%-17.9% versus 9.1-10.0%) it is worth noting the improved quality of analytical results due only to the data treatment.

Along with wCal-IS, IPD stands out in comparison with more extensively used methods such as Cal-IS. Furthermore, IPD also provided combined uncertainties below 8.4% in all the concentration range with the advantage of reduced time analysis, since no calibration curve procedure had to be performed.

Again, results showed that Cal-IS performs poorly at low concentrations, being the worse method for the whole concentration range studied. In comparison, wCal-IS provided constant values of combined uncertainty along the concentration range although higher than IPD, which produced the lowest combined uncertainties of the three methods at any concentration assayed. This is in accordance with the high metrological quality of analytical results provided by isotope dilution mass spectrometry determinations[21].

Finally, full uncertainty budgets were obtained for the three selected quantification methods according to the Kragten approach (Table S.6 in supplementary material).

In the case of both wCal-IS and Cal-IS methods, the same 6 parameters were considered, including: intercept and slope of the linear regression, measurement of the area ratio in the sample (between natural

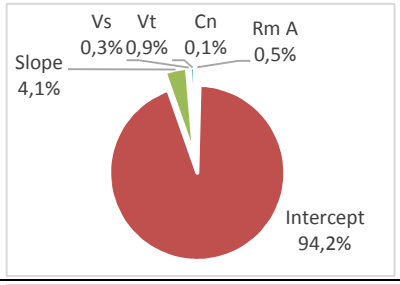
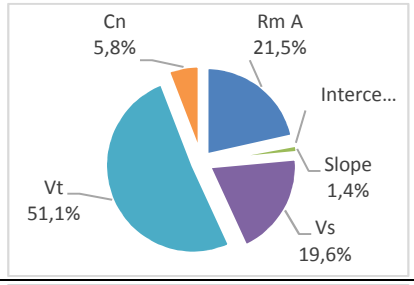
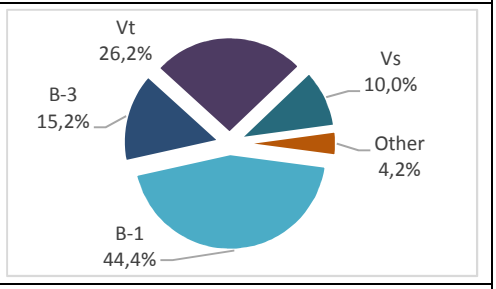
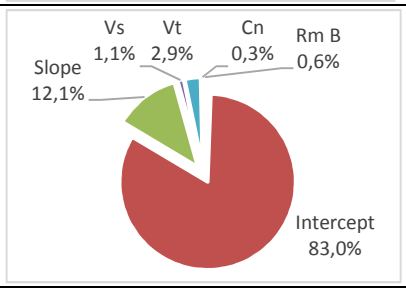
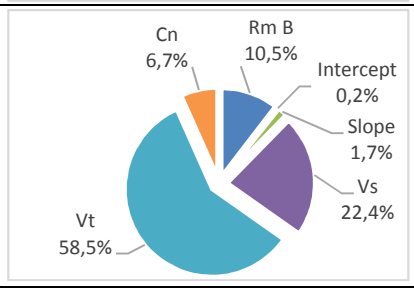
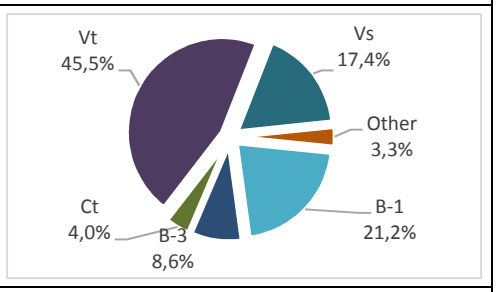
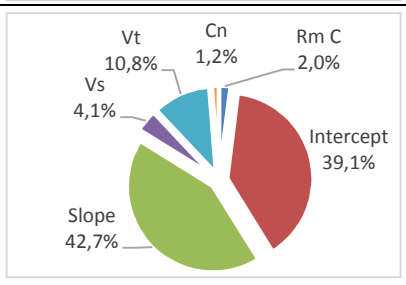
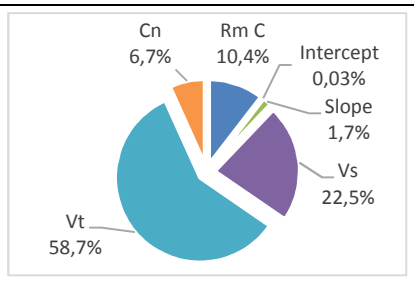
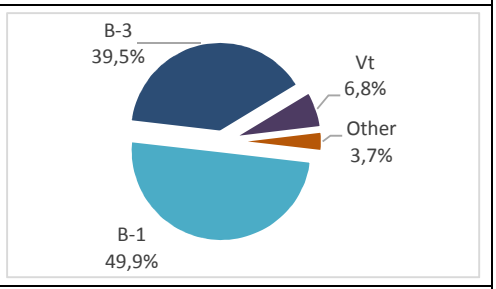
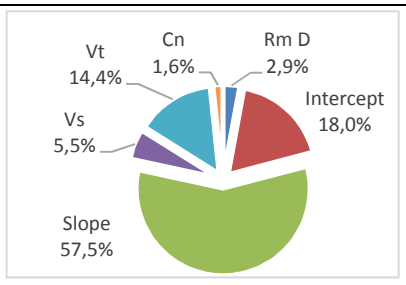
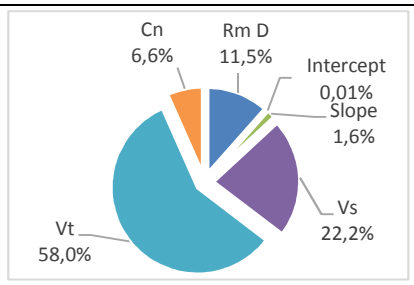
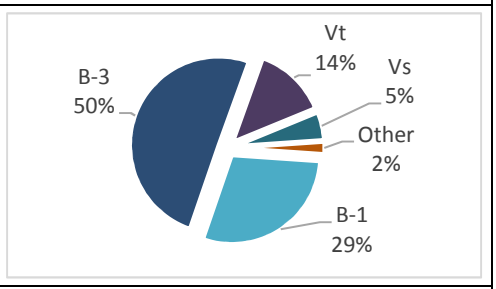
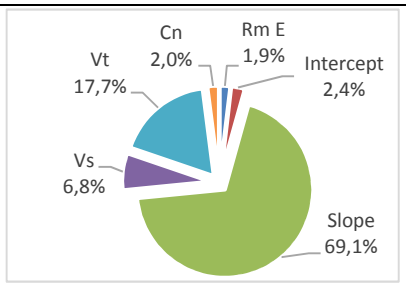
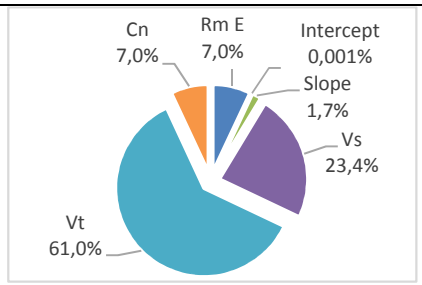
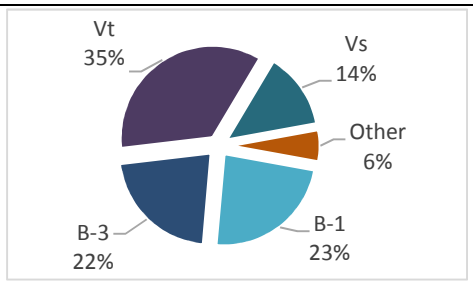
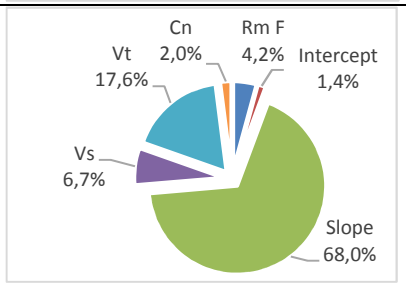
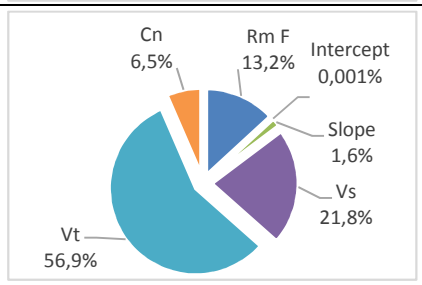
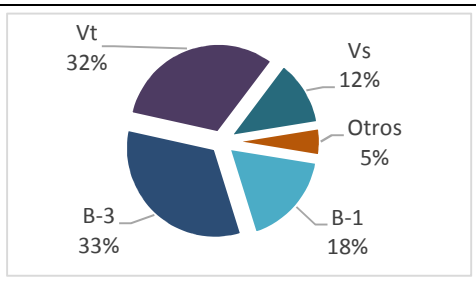
and labelled compound chromatographic peak areas, *Rm*), volume of sample (*Vs*), volume of internal standard (*Vt*) and concentration of the natural standard (*Cn*). Calculations of the contribution of each parameter to total procedure uncertainty were carried out for the five replicates and the average values were obtained.

As it can be seen in Figure 2, in the case of Cal-IS, uncertainty contribution coming from the intercept of the regression is predominant at low concentrations (Sample A) while at high concentrations (Sample F) slope is the highest contributor to final method uncertainty. Thus, uncertainty for a Cal-IS method will hardly improve experimentally. Probably, an alternative way to correct bias and its associated uncertainty at low concentrations could be the use of a single external calibration point forced through the origin, an approach not tested in the present work. In contrast, when using weighted calibration, the major contributors to uncertainty were the measurement of sample and internal standard volumes. Thus, one way to easily reduce uncertainty could be consider the mass instead of volumes.

On the other hand, the parameters considered for uncertainty calculations in IPD quantification were the following: determination of abundances of the natural testosterone (*natT-1, natT-2*) and $^{13}C_2$-testosterone (*labT-1, labT-2*) transitions, measurement of those transitions in the sample blend (*B-1, B-2, B-3, B-4*), volume of sample (*Vs*) and volume and concentration of $^{13}C_2$-testosterone standard added (*Vt, Ct*).

As it might be expected for an isotope-dilution determination[21], one of the most important parameters in IPD was the volume of labelled compound added to perform the quantification. As said above, this contribution to uncertainty could be minimized by weighting the amount of solution added. Moreover, the measurement of relative abundances in the sample blend, especially the most abundant transitions of natural (*B-1*: 289 > 97) and labelled compounds (*B-3*: 291 > 99), contributed significantly to the final uncertainty with relative magnitudes from 8.6% to 50%.

340

Sample	Non-weighted calibration	Weighted calibration	IPD
A			
B			
C			
D			
E			
F			

341

342 Figure 2. Uncertainty budgets for the quantification methods assayed.

343

CONCLUSIONS

In this work, three analytical approaches for the determination of testosterone in urine have been compared from an uncertainty evaluation point of view.

Firstly, method uncertainty derived from the procedure itself has been evaluated at our laboratory by applying weighted and non-weighted calibration with internal standard and IPD quantifications to 6 synthetic urine samples, composed of mixed human urine samples, in five different weeks. Inter-day combined uncertainties for each sample and method were obtained by Nordtest calculation method and showed similar values for weighted calibration and IPD, below or equal to 10% in all cases, while non-weighted calibration yielded uncertainties ranging from 11.2% to 17.9%.

Secondly, an inter-laboratory experiment was carried out in order to set a reference value for the samples and to further evaluate inter-laboratory RSD of these three methods. Similarly to the intra-laboratory experiment, non-weighted calibration presented much higher uncertainty at low concentrations (43%) than at medium and high concentrations (12%-24%), where it showed a better performance than weighted calibration (18%-21% along all the range). In contrast, the combined uncertainty associated with IPD method was lower than the other two in all 6 samples, ranging from 7.8% to 13%.

In addition, Kragten method was applied to intra-laboratory data to obtain the uncertainty budgets for the considered quantification methods. Thus, linear regression parameters –slope and intercept– were found to be the major contributors to uncertainty in non-weighted calibration, varying along the concentration range. In contrast, weighted calibration and IPD methods were more stable in terms of relative contributions to procedure uncertainty.

Hence, it has been demonstrated that weighted calibration might be more precise than classical calibration with internal standard, providing similar uncertainties and standard deviations than isotope dilution methodologies in intra-laboratory reproducibility studies. Moreover, the present IPD methodology yielded lower inter-laboratory variability and thus, higher metrological quality of the analytical results are expected.

The results presented in this work for testosterone as a model compound, together with the benefits of reduced time analysis and matrix effect corrections provided by IDMS-based methodologies, highlights IPD as a rapid, robust and reliable method. Thus, taking into account the lower uncertainty of the present analytical approach, IPD is shown as a promising alternative to improve longitudinal fluctuations in steroid profiling.

ACKNOWLEDGEMENTS

The authors acknowledge financial support from the Generalitat Valenciana (Research group of excellence Prometeo II 2014/023 and Collaborative Research on Environment and Food Safety ISIC/2012/016), as well as University Jaume I for project PB1-1B2013-55. Finally, the authors are grateful to the Serveis Centrals d'Instrumentació Científica (SCIC) of University Jaume I for using Acquity and TQD instruments.

383 **REFERENCES**

384 [1] F. Gosetti, E. Mazzucco, M.C. Gennaro, E. Marengo, Ultra high performance liquid chromatography
385 tandem mass spectrometry determination and profiling of prohibited steroids in human biological
386 matrices. A review, J. Chromatogr. B. 927 (2013) 22–36. doi:10.1016/j.jchromb.2012.12.003.

387 [2] F. Botrè, New and old challenges of sports drug testing, J. Mass Spectrom. 43 (2008) 903–907.
388 doi:10.1002/jms.1455.

389 [3] A. Jarek, K. Kowalczyk, P. Chołbiński, K. Chajewska, E. Turek-Lepa, A. Pokrywka, E. Bulska, D.
390 Kwiatkowska, Analytical procedure for steroid profiling valid for Athlete Biological Passport, Chem.
391 Pap. 69 (2015) 254–261. doi:10.1515/chempap-2015-0030.

392 [4] C. Saudan, Testosterone and doping control, Br. J. Sports Med. 40 (2006) i21–i24.
393 doi:10.1136/bjsm.2006.027482.

394 [5] F. Badoud, D. Guillarme, J. Boccard, E. Grata, M. Saugy, S. Rudaz, J.-L. Veuthey, Analytical aspects in
395 doping control: Challenges and perspectives, Forensic Sci. Int. 213 (2011) 49–61.
396 doi:10.1016/j.forsciint.2011.07.024.

397 [6] R. Nicoli, D. Guillarme, N. Leuenberger, N. Baume, N. Robinson, M. Saugy, J.-L. Veuthey, Analytical
398 Strategies for Doping Control Purposes: Needs, Challenges, and Perspectives, Anal. Chem. 88 (2016)
399 508–523. doi:10.1021/acs.analchem.5b03994.

400 [7] G.J. Trout, R. Kazlauskas, Sports drug testing - an analyst's perspective, Chem. Soc. Rev. 33 (2004) 1.
401 doi:10.1039/b201476a.

402 [8] World Anti-Doping Agency, WADA Technical Document – TD2016EAAS, (2016) 1–16.

403 [9] M. Thevis, A. Thomas, V. Pop, W. Schänzer, Ultrahigh pressure liquid chromatography–(tandem)
404 mass spectrometry in human sports drug testing: Possibilities and limitations, J. Chromatogr. A. 1292
405 (2013) 38–50. doi:10.1016/j.chroma.2012.12.048.

406 [10] P.J. Taylor, Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography–
407 electrospray–tandem mass spectrometry, Clin. Biochem. 38 (2005) 328–334.
408 doi:10.1016/j.clinbiochem.2004.11.007.

409 [11] A. Furey, M. Moriarty, V. Bane, B. Kinsella, M. Lehane, Ion suppression; A critical review on causes,
410 evaluation, prevention and applications, Talanta. 115 (2013) 104–122.
411 doi:10.1016/j.talanta.2013.03.048.

412 [12] H. Trufelli, P. Palma, G. Famiglini, A. Cappiello, An overview of matrix effects in liquid
413 chromatography-mass spectrometry, Mass Spectrom. Rev. 30 (2011) 491–509.
414 doi:10.1002/mas.20298.

415 [13] A. González-Antuña, P. Rodríguez-González, G. Centineo, J.I. García Alonso, Evaluation of minimal
416 13C-labelling for stable isotope dilution in organic analysis, Analyst. 135 (2010) 953.
417 doi:10.1039/b924432h.

418 [14] N. Fabregat-Cabello, J. Pitarch-Motellón, J. V. Sancho, M. Ibáñez, A.F. Roig-Navarro, Method
419 development and validation for the determination of selected endocrine disrupting compounds by
420 liquid chromatography mass spectrometry and isotope pattern deconvolution in water samples.

421 Comparison of two extraction techniques, Anal. Methods. 8 (2016) 2895–2903.
422 doi:10.1039/C6AY00221H.

423 [15] N. Fabregat-Cabello, P. Zomer, J.V. Sancho, A.F. Roig-Navarro, H.G.J. Mol, Comparison of approaches
424 to deal with matrix effects in LC-MS/MS based determinations of mycotoxins in food and feed, World
425 Mycotoxin J. 9 (2016) 149–161. doi:10.3920/WMJ2014.1872.

426 [16] N. Fabregat-Cabello, J. V Sancho, A. Vidal, F. V González, A.F. Roig-Navarro, Development and
427 validation of a liquid chromatography isotope dilution mass spectrometry method for the reliable
428 quantification of alkylphenols in environmental water samples by isotope pattern deconvolution., J.
429 Chromatogr. A. 1328 (2014) 43–51. doi:10.1016/j.chroma.2013.12.077.

430 [17] E. Spirito, F. Botre, The role of measurement uncertainty in doping analysis, Int. J. Risk Assess.
431 Manag. 5 (2005) 374. doi:10.1504/IJRAM.2005.007178.

432 [18] O. Barroso, J. Miller, A. Squirrell, S. Westwood, Measurement uncertainty in anti-doping quantitative
433 analysis for prohibited threshold substances, Bioanalysis. 4 (2012) 1653–1665.
434 doi:10.4155/bio.12.127.

435 [19] J.I. García-Alonso, P. Rodríguez-González, Uncertainty evaluation in IDMS, in: Isot. Dilution Mass
436 Spectrom., RSC Publishing, 2013: pp. 379–433

437 [20] L. Ramaley, L.C. Herrera, Software for the calculation of isotope patterns in tandem mass
438 spectrometry, Rapid Commun. Mass Spectrom. 22 (2008) 2707–2714. doi:10.1002/rcm.3668.

439 [21] J.I. García-Alonso, P. Rodríguez-González, IDMS procedures and calculation methods, in: Isot. Dilution
440 Mass Spectrom., RSC Publishing, 2013: pp. 41–76

441 [22] P. Rodríguez-González, J.I. García Alonso, Recent advances in isotope dilution analysis for elemental
442 speciation, J. Anal. At. Spectrom. 25 (2010) 239. doi:10.1039/b924261a.

443 [23] Á. Castillo, E. Gracia-Lor, A.F. Roig-Navarro, J.V. Sancho, P. Rodríguez-González, J.I.G. Alonso, Isotope
444 pattern deconvolution-tandem mass spectrometry for the determination and confirmation of
445 diclofenac in wastewaters, Anal. Chim. Acta. 765 (2013) 77–85. doi:10.1016/j.aca.2012.11.033.

446 [24] World Anti-Doping Agency, Technical Document-TD2014DL Decision Limits for the Quantification of
447 Threshold Substances, (n.d.).

448 [25] Nordtest, Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories,
449 (2012).

450 [26] J. Kragten, Tutorial review. Calculating standard deviations and confidence intervals with a
451 universally applicable spreadsheet technique, Analyst. 119 (1994) 2161. doi:10.1039/an9941902161.

452