Experimental and statistical protocol for the effective validation of chromatographic analytical methods

(Article begins on next page)

Please use this MethodsX article template when submitting your Method as a companion paper
to a research article to _MethodsX_

A MethodsX article focuses on the technical aspect of your work, and provides evidence of the efficiency of your method/comparison with pre- existing protocols. This should be immediately evident to the reader

---

- _Please fill in the template below and delete all instruction text above and below before submitting._
- _All sections indicated by * are compulsory._
- _NOTE: - submitting as a companion paper to a research article, zip all files relevant to the MethodsX submission, including the graphical abstract, into a single .zip file and upload as a "Method article" item._

- _If you have any questions concerning the submission of your MethodsX paper, please contact the journal at mexjm@elsevier.com_

Meta-Data (Required for the transfer of your article to MethodsX – will not be typeset)

*Title:	_Experimental and statistical protocol for the effective validation of chromatographic analytical methods_
*Authors:	_Eleonora Amante [a,b], Eugenio Alladio [a,b], Cristina Bozzolino [a], Fabrizio Seganti [b], Alberto Salomone [a,b], Marco Vincenti [a,b], Brigitte Desharnais [c]_
*Affiliations:	_[a] Dipartimento di Chimica, Università degli Studi di Torino, Turin, Italy_ _[b] Centro Regionale Antidoping e di Tossicologia "A. Bertinaria", Orbassano (Turin), Italy_ _[c] Laboratoire de sciences judiciaires et de médecine légale, Montreal, Quebec, Canada_
*Contact email:	_eamante@unito.it_
*Co-authors:	_Eugenio Alladio: ealladio@unito.it_ _Cristina Bozzolino:_ _cristina.bozzolino@hotmail.it_ _Fabrizio Seganti:_

	fabrizio.seganti@antidoping.piemonte.it *Alberto Salomone: alberto.salomone@unito.it* *Marco Vincenti: marco.vincenti@unito.it* *Brigitte Desharnais:* *brigitte.desharnais@msp.gouv.qc.ca*
*Keywords: *At least 3 keywords.* • *There is no limit on the no. of keywords you can list.* • *Please remember that effective keywords should not repeat words appearing in your title, and should be neither too general nor too narrow.*	*Chromatographic method* *Validation protocol* *multiresidual analysis* *GC-MS*
*SECTION: • *Agricultural and Biological Sciences* • *Biochemistry, Genetics and Molecular Biology* • *Chemical Engineering* • *Chemistry* • *Computer Science* • *Earth and Planetary Sciences* • *Energy* • *Engineering* • *Environmental Science* • *Immunology and Microbiology* • *Materials Science* • *Mathematics* • *Medicine and Dentistry* • *Neuroscience* • *Pharmacology, Toxicology and Pharmaceutical Science* • *Physics and Astronomy* • *Psychology* • *Social Sciences* • *Veterinary Science and Veterinary Medicine*	*Chemistry*

Method Article

Title:

Authors:

Eugenio Alladio [a,b#], Eleonora Amante [a,b#], Cristina Bozzolino [a], Fabrizio Seganti[.b], Alberto Salomone [a,b], Marco Vincenti [a,b], Brigitte Desharnais [c]*

Affiliations:

[a] *Dipartimento di Chimica, Università degli Studi di Torino, Turin, Italy*

[b] *Centro Regionale Antidoping e di Tossicologia "A. Bertinaria", Orbassano (Turin), Italy*

[c] *Laboratoire de sciences judiciaires et de médecine légale, Montreal, Quebec, Canada*

Contact email:

* eamante@unito.it

# The authors equally contributed to this work.

Abstract

- The validation of analytical methods is of crucial importance in several fields of application. A new protocol for the validation of chromatographic methods has been proposed, with a focus on the calibration study, by the evaluation of heteroscedasticity, comparison of several weights and linear/quadratic calibration curves. The operating protocol foresees the repetition of three calibration curves in three different days, providing a total amount of nine replicates. Such a structured design allows to use the same experiments to determine several parameters which are traditionally computed from dedicated experiments: intra- and inter-day accuracy and precision, limit of detection, specificity, selectivity, ion abundance repeatability and carry over.

- Few further experiments are necessary to evaluate the retention time repeatability, matrix effect and extraction recovery.

- The overall protocol is described in a parallel paper, where the case of a multi-targeted gas chromatography – mass spectrometry (GC-MS) method for the determination of androgens in human urine is in-depth discussed. The purpose of this paper is to report the details about the GC-MS separation and detection of the target analytes, and to provide the mathematical formulas needed to perform the validation of the principal parameters.

Max 200 words, include up to 3 bullet points.
- *Remember that the abstract is what readers see first in electronic abstracting & indexing services.*
- *This is the advertisement of your article. Make it interesting, and easy to be understood.*
- *Be accurate and specific, keep it as brief as possible*
- *Add up to 3 bullet points, highlighting the customization rather than the steps of the procedure. Highlight how it differs/which advantage it offers over pre-existing methods.*

Specifications Table

Subject area	*Chemistry*
More specific subject area	*Analytical Chemistry*
Method name	*Effective validation protocol for chromatography – mass spectrometry analytical methods*
Name and reference of original method	*Not applicable*
Resource availability	*If applicable, include links to resources necessary to reproduce the method (e.g. data, software, hardware, reagent)*

Method details

This paper accompanies the paper entitled "*Effective validation of chromatographic analytical methods: the illustrative case of androgenic steroids*", which presents a new, systematic validation protocol for chromatographic analytical methods. As case study, the fully validation of a multiresidual GC-MS method for the detection of androgens is human urine is discussed. The details related to the separation and acquisition methods are reported in this paper; specifically, the oven temperature program of the gas chromatograph is reported in Figure 1, together with the typical total ion current (TIC) profile of a real urine sample. Moreover, details about the mass spectrometer (MS) detection of the 18 target compounds (i.e. retention time, quantifier and monitored ions) plus the molecular weight after trimethylsilyl derivation are in Table 1.

Furthermore, the validation protocol is described in the Experimental Design Section, and all the parameters (homoscedasticity evaluation, linearity tests such as ANOVA, Mandel's test and Lack of Fit, limit of detection, intra- and inter-day accuracy and precision, matrix effect, extraction recovery) are defined, together with the equations for their computations.

Experimental Design, Materials, and Methods

Analytical method

Samples pre-treatment

The sample preparation involved the fortification of 6 mL of urine with testosterone-D$_3$ and 17α-methyltestosterone at the final concentration of 25 ng/mL and 125 ng/mL, respectively. The pH was then adjusted to a value between 6.8 and 7.4 by adding 2 mL phosphate buffer 0.1 M and drop(s) of NaOH 1 M, if necessary. A volume of β-glucuronidase solution corresponding to 83 units was added and then the mixture was incubated at 58 °C for 1 hour. After cooling at room temperature, 2 mL carbonate buffer 0.1 M was added to the aqueous solution, together with drop(s) of NaOH 1 M, until the final pH = 9 was reached. Then, liquid-liquid extraction (LLE) was performed with 10 mL of TBME; the samples were shaken in a multi-mixer for 10 minutes, centrifuged at 6.24 g for 5 minutes and the organic supernatant was transferred into a glass tube. The extracts were subsequently dried under a nitrogen flow at 70 °C. After addition of 50 µL derivatizing solution (MSTFA/NH4I/dithioerythritol – 1,000:2:4 v/w/w), the reaction was allowed to proceed at 70 °C for 30 minutes. The resulting solutions were transferred into conical vials and a 1 µL aliquot was injected by autosampler into the GC-MS working in the splitless mode. Mix I and II had distinct calibration ranges (Table 1), selected on the basis of the expected physiological concentrations, as reported in literature [1,2].

	Target analyte	CAS number	Internal standard
Mix I	**5β-androstan-13,17-dione**	1229-12-5	Testosterone-D$_3$
	5α-androstane-3α,17β-diol (5α-adiol)	1852-53-5	Testosterone-D$_3$
	5β-androstane-3α,17β-diol (5β-adiol)	1851-23-6	Testosterone-D$_3$
	dehydroepiandrosterone (DHEA)	53-43-0	Testosterone-D$_3$
	5-androsten-3,17-diol	512-17-5	Testosterone-D$_3$
	epitestosterone (E)	481-30-1	Testosterone-D$_3$
	4,6-androstadien-3,17-dione (6-D)	633-34-1	Testosterone-D$_3$
	dihydrotestosterone (DHT)	521-18-6	Testosterone-D$_3$
	4-androsten-3,17-dione	63-05-8	Testosterone-D$_3$
	Δ6-testosterone	2484-30-2	Testosterone-D$_3$
	testosterone (T)	58-22-0	Testosterone-D$_3$
	7α-hydroxytestosterone	62-83-9	Testosterone-D$_3$
	7β-hydroxy-dehydroepiandrosterone (7β-OH-DHEA)	2487-48-1	Testosterone-D$_3$
	formestane	566-48-3	Testosterone-D$_3$
	4-hydroxytestosterone	2141-17-5	Testosterone-D$_3$
	16α-hydroxyandrosten-3,17-dione	63-02-5	Testosterone-D$_3$
Mix II	**androsterone (Andro)**	53-41-8	17α-methyl-testosterone
	etiocholanolone (Etio)	53-42-9	17α-methyl-testosterone

Calibration level	1	2	3	4	5	6
Mix I *(ng/mL)*	2	5	10	25	50	125
Mix II *(ng/mL)*	100	200	500	1000	1500	2250

Table 1. List of the analytes included in MIX I and MIX II, with the relative CAS number and the internal standard used for their quantitation. The concentrations at the different calibration levels are also reported.

GC-MS separation and detection

The GC-MS method optimization was the subject of another study [2]. The GC separation was performed using an Agilent 6890N instrument (Agilent Technologies, Milan, Italy) equipped with a J&W Scientific HP-1, 17 m x 0.2 mm (i.d.) x 0.11 mm (f.t.) capillary column. Helium was employed as the carrier gas at a constant pressure of 18.5 psi. The temperature program of the GC oven was set as follows: initial temperature equal to 120 °C, then a 70 °C/min heating rate was applied until the temperature of 177 °C was reached. Subsequently, the temperature was raised to 236 °C with a 5 °C/min gradient. A final heating rate of 30 °C/min allowed to rise the temperature of 315 °C, which was hold for 3 minutes. The GC injector and transfer line were maintained at 280 °C. The temperature program is reported in Figure 1 (blue line).
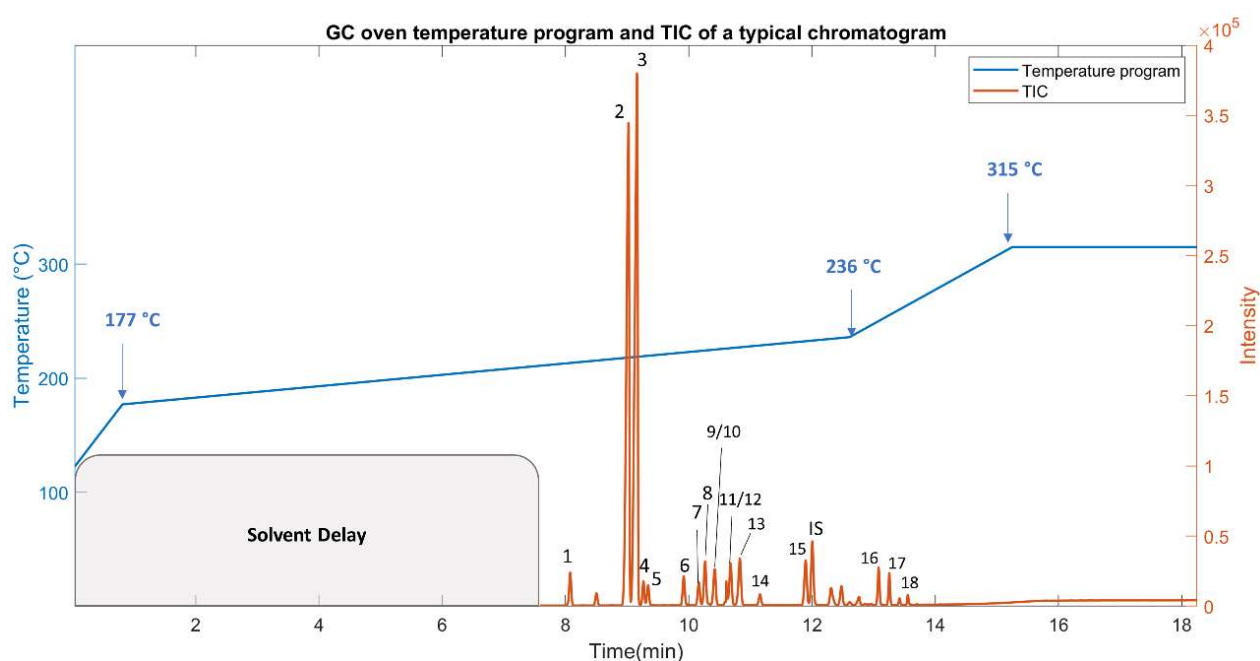
Figure 1. Temperature program of the GC oven (blue line) and typical chromatographic profile (orange line). Coded target analytes are: (1) 5β-androstan-3,17-dione, (2) A, (3) Etio, (4) 5α-adiol, (5) 5β-adiol, (6)DHEA, (7) 5-androsten-3,17-diol, (8) E, (9) 4,6-androstadien-3,17-dione, (10) DHT, (11) 4-androsten-3,17-dione, (12) Δ6-testosterone, (13) testosterone + testosterone-D3, (14) 7α-hydroxytestosterone, (IS) 17-methyl-testosterone, (15) 7β-OH-DHEA, (16) Formestane, (17) 4-hydroxytestosterone, (18) 16α-hydroxyandrosten-3,17-dione.

The trimethylsilyl derivatives of the analytes were ionized and fragmented in EI at 70 eV using an Agilent 5975 inert mass-selective detector (Agilent Technologies, Milan, Italy). The MS was operated in the selected ion monitoring mode and three diagnostic ions for each analyte were monitored with dwell times of 20 - 50 ms. The retention times and characteristic ions monitored are reported in Table 2, together with the molecular weight of the derivatized compounds. In Figure 1 (orange line and Arabic numbers) is reported the typical Total ion current (TIC) profile of a spiked urine sample.

Target analyte	Molecular weight (after derivatisation)	T_R (min)	Ion 1 (m/z)	Ion 2 (m/z)	Ion 3 (m/z)
5β-androstan-13,17-dione	432	8.07	**275**	290	417
Androsterone (Andro)	434	9.02	**419**	434	329
Etiocholanolone (Etio)	434	9.16	**419**	329	434
5α-androstane-3α,17β-diol (5α-adiol)	436	9.26	**241**	215	256
5β-androstane-3α,17β-diol (5β-adiol)	436	9.34	**256**	241	215
Dehydroepiandrosterone (DHEA)	432	9.92	**432**	417	327
5-androsten-3,17-diol	434	10.16	**239**	215	254
Epitestosterone (E)	432	10.27	**432**	327	417

4,6-androstadien-3,17-dione	428	10.42	**428**	413	323
Dihydrotestosterone (DHT)	434	10.43	**434**	405	419
4-androsten-3,17-dione	430	10.61	**430**	415	325
Δ6-testosterone	430	10.68	**430**	415	325
Testosterone (T)	432	10.82	**432**	417	327
7α-hydroxytestosterone	520	11.15	**520**	431	296
7β-hydroxydehydroepiandrosterone (7β-OH-DHEA)	520	11.90	**430**	415	325
Formestane	518	13.08	**518**	430	503
4-hydroxytestosterone	520	13.25	**520**	505	447
16α-hydroxyandrosten-3,17-dione	518	13.55	**503**	518	430
testosterone-d₃	435	10.82	**435**	420	
17α-methyltestosterone	446	12.00	**301**	446	

Table 2. Molecular weight, retention time and characteristic ions of the 17 monitored steroids and the two internal standards.

Validation protocol

The validation protocol is in-depth described in the parallel paper [REF]. Briefly, nine replicates of the calibration curve are analyzed in three different days (three replicates/die). This peculiar experimental design allowed the simultaneous evaluation of several parameters, which are typically evaluated performing dedicated experiments, resulting in expensive and timewasting protocols. Among these, a particular focus was put on the study of the calibration curve, with tests of homoscedasticity, quadraticity, ANOVA, Lack of Fit and goodness of the back

calculation. The calibration curves were also used for the evaluation of the limit of detection (LOD, by Hubaux and Vos' approach) and intra- and inter-day accuracy and precision. Furthermore, ion abundance repeatability, selectivity, specificity and carry over were studied employing the same experiments. Lastly, few further experiments were performed to determine matrix effect and extraction recovery.

The principal equations employed are reported below.

Nomenclature

In this article, the calibration levels are indicated as *1,2,...i,....k* and the replicates as *1,2,...j,....l* and the total number of samples analyzed is *k x l = n*.

Computation of the calibration model

Test for heteroscedasticity

The homoscedasticity was tested twice, e.g. using a partial F-Test integrated in the R routine (Equation 1) developed by Desharnais *et al.* [3] and the Levene equation (Equation 2) [4]. In the first case, the presence of heteroscedasticity was investigated using a unilateral F-test for the calculation of the probability that the variance of measurements at the upper limit of quantification (ULOQ) was equal to or smaller than the variance of measurements at the lower limits of quantification (LLOQ). The Rstudio function used for the computation is

$$var.test(Measurements_{LLOQ}, Measurements_{ULOQ}, alternative = "less") \qquad (1)$$

Unlike the unilateral F-test described above, the Levene test was applied on all the calibration levels. It is a robust alternative to the F-test and was used to confirm the results obtained with the calibration routine. The equation of Levene test is the following:

$$W = \frac{(n-k)}{(k-1)} \times \frac{\sum_{i=1}^{k} l(\bar{Z}_i - \bar{Z})^2}{\sum_{i=1}^{k} \sum_{j=1}^{l}(Z_{ij} - \bar{Z}_i)^2} \qquad (2)$$

With

$$Z_{ij} = |y_{ij} - \bar{y}_i|$$

k is the number of calibration levels tested, $\bar{Z}_i$ is the average of all the Z_{ij} of a calibration level and $\bar{Z}$ is the average of all the Z_{ij}, in the original version of the test, or their median, from the Brown-Forsythe modification, which is more robust towards heavy-tailed distributions [5]. The RStudio function *levene.test* was used to perform the calculations (in the Brown-Forsythe version).

The W statistics can be compared to an F distribution with {(k-1),(n-k)} degrees of freedom. If the p-value is smaller than the α level of significance chosen (in our case, 0.05), then the variances are considered as significantly different, i.e. the data are heteroscedastic. If p > α, the data is consistent with an equality of variances.

Partial F-test for the quadratic term

The Partial F-test is a hypothesis test which relies on comparing the sum of squares of the regression to the mean square of residuals (Equation 3):

$$F_{exp} = \frac{SS_{reg,Q} - SS_{reg,L}}{\left(\frac{SS_{res,Q}}{n-3}\right)} \qquad (3)$$

Where $SS_{reg,Q}$ and $SS_{reg,L}$ are the sum of squares of the regression in the quadratic and linear models, respectively (Equation 7). SS_{res} is the sum of mean squares in residuals (Equation 8).

$$SS_{reg} = \sum_{i=1}^{k} w_i \times l \times (\hat{y}_i - \bar{y}_{ij})^2 \qquad (4)$$

$$SS_{res} = \sum_{i=1}^{k} \sum_{j=1}^{l} w_i \times (y_{ij} - y_i)^2 \qquad (5)$$

The p-value associated with F_{exp} can be found using the RStudio command *1-pf(F_{calc},1,(n-3))*. A $p < 0.05$ denotes a significant improvement in the model fit brought by the use of a quadratic model.

Analysis of Variance - Lack of Fit (ANOVA-LoF) to verify the goodness of the calibration model

The ANOVA-LoF hypothesis test is used to evaluate the fit of data-points with the final calibration model. The null-hypothesis is that there is no lack of fit and the F is computed as follows (equation 6):

$$F_{LoF} = \frac{\dfrac{SS_{LoF}}{DoF_{LoF}}}{\dfrac{SS_{Pure}}{DoF_{Pure}}} = \frac{\dfrac{\sum_{i=1}^{k} \sum_{j=1}^{l} w_i \times (\bar{y}_i - \hat{y}_i)^2}{k - q}}{\dfrac{\sum_{i=1}^{k} \sum_{j=1}^{l} w_i \times (y_{ij} - y_i)^2}{n - k}} \qquad (6)$$

It is important to underline that this test is very sensitive to experimental design, in particular the number of replicates and/or the number of calibration levels. Hence, if accuracy and precision are within the limits of acceptability, it is possible to ignore the outcome of this test.

Limit of Detection (LOD)

The limit of detection is the lower concentration detectable with the specified analytical method. It can be evaluate using several different approaches; here, we propose the Hubaux and Vos' computation[6].

The approach relies on five hypotheses:

1. The standards are independent

2. The contents of the standards are accurately known

3. The observed signals have a gaussian distribution

4. A linear regression model is adequate for the data at hand

5. The variance of the error is constant (i.e. homoscedastic data).

Assuming that the first three prerequisites are met, it is necessary to focus on numbers 4 and 5, which are not necessarily respected. When linearity is not respected, it is possible to reduce the calibration range excluding the upper calibration levels, in order to exclude the quadraticity.

If the homoscedasticity is not respected, the weights need to be introduced into the Hubaux and Vos equation (equations 6-10):

$$X_{LOD} = t_{(0.05, n-2)} \times s_{\hat{y}0} \tag{6}$$

Where t is the Student's test, value at 0.05 confidence limit and n-2 degrees of freedom, and s_{y0} is equal to

$$s_{\hat{y}0} = S_{y/x} \sqrt{1 + \frac{1}{\sum_{i=1}^{k} l \times w_i} + \frac{(-\bar{x}_w)^2}{\sum_{i=1}^{k} l \times w_i (x_i - \bar{x}_w)^2}} \tag{7}$$

$S_{y/x}$ and x_w are, respectively:

$$S_{y/x} = \sqrt{\frac{\sum_{i=1}^{k} \sum_{j=1}^{l} w_i \times (y_i - y_{ij})^2}{n-2}} \tag{8}$$

$$\bar{x}_w = \frac{\sum_{i=1}^{k} \sum_{j=1}^{l} w_i x_i}{\sum_{i=1}^{k} \sum_{j=1}^{l} w_i} \tag{9}$$

And, finally,

$$Y_c = bX_c + a \tag{10}$$

a is the intercept of the calibration curve and b the slope of the calibration curve.

Once the concentration of analyte constituting the LOD is mathematically obtained, an experimental verification is needed. It consists in the fortification of blank matrix at the computed X_{LOD} and the measurement of the Signal-to-Noise, which has to be higher than 3.

Accuracy

The accuracy is a measure of the closeness of a measured value to the actual value. The three replicates measured in each validation day allow the computation of the intra-day accuracy, and the 12-days timeframe of the overall validation procedure allows the evaluation of the inter-day accuracy. The two computations are performed employing the R routine developed by Desharnais et al.[3,7] following the operating scheme presented here [REF articolo]. The method's accuracy is expressed in terms of bias%, which it is measured as follows:

$$bias_\% = \left(1 - \frac{x_{real}}{\bar{x}_{exp}}\right) \times 100 \tag{11}$$

Where x_{real} is the spiked concentration and $\bar{x}_{exp}$ is the experimental result.

Precision

The precision is the reproducibility of a measurement, i.e. describes how close are the replicates.

It is expressed as coefficient of variance and computed as follows:

$$CV_\% = \frac{\frac{\sum_{j=1}^{J}(x_{exp} - \bar{x})^2}{J-1}}{\bar{x}} \times 100 \tag{12}$$

Where J is the number of replicates, x_{exp} is the experimental result of the j-replicate and $\bar{x}$ is the mean result.

Matrix effect (ME)

To evaluate the ME, bi-distilled water and synthetic urine are spiked, after the extraction step, at the desired concentration (typically, three concentration levels are tested, i.e. low, middle and high). The ME is provided by the ratio of the means of the replicates (minimum of three):

$$ME_\% = \frac{\overline{\left(A_S/A_{IS}\right)}_w}{\left(A_S/A_{IS}\right)_u} \times 100 \tag{13}$$

Where A_S and A_{IS} are the area of the standard and the internal standard, respectively; w indicates bi-distilled water and u synthetic urine. Values between 85% and 115% are considered acceptable.

Extraction Recovery (ER)

The ER is evaluated comparing the results obtained spiking the standards and internal standards before and after the extraction procedure. The number of replicates and the concentration levels usually tested are the same reported in the ME description. The formula is the following:

$$ER_\% = \frac{\overline{\left(A_S/A_{IS}\right)}_{before}}{\left(A_S/A_{IS}\right)_{after}} \times 100 \tag{14}$$

Where A_S and A_{IS} are defined as above, *before* and *after* are the samples spiked before and after the extraction, respectively. Again, values between 85% and 115% are considered acceptable.

*References: *[Include at least one reference, to the original publication of the method you customized.]*

[1]   P. Van Renterghem, P. Van Eenoo, H. Geyer, W. Schänzer, F.T. Delbeke, Reference ranges for urinary concentrations and ratios of endogenous steroids, which can be used as markers for steroid misuse, in a Caucasian population of athletes, Steroids. 75 (2010) 154–163. doi:10.1016/j.steroids.2009.11.008.

[2]   E. Amante, E. Alladio, A. Salomone, M. Vincenti, F. Marini, G. Alleva, S. De Luca, F. Porpiglia, Correlation between chronological and physiological age of males from their multivariate urinary endogenous steroid profile and prostatic carcinoma-induced deviation, Steroids. 139 (2018) 10–17. doi:10.1016/j.steroids.2018.09.007.

[3]   B. Desharnais, F. Camirand-Lemyre, P. Mireault, C.D. Skinner, Procedure for the selection and validation of a calibration model II-theoretical basis, J. Anal. Toxicol. 41 (2017) 269–276. doi:10.1093/jat/bkx002.

[4]   Levene, Robust tests for equality of variances, in: I. Olkin, H. Hotelling (Eds.), Contrib. to Probab. Stat. Essays Honor Harold Hotell., Stanford University Press, 1960: pp. 278–292.

[5]   M.B. Brown, A.B. Forsythe, Robust tests for the equality of variances, J. Am. Stat. Assoc. 69 (1974) 364–367.

[6]   A. Hubaux, G. Vos, Decision and Detection limits for linear Calibration Curves, Anal. Chem. 42 (1970) 849–855. doi:10.1021/ac60290a013.

[7]   B. Desharnais, F. Camirand-lemyre, P. Mireault, C.D. Skinner, Procedure for the Selection and Validation of a Calibration Model I — Description and Application, J. Anal. Toxicol. 41 (2017) 261–268. doi:10.1093/jat/bkx001.