

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Effective validation of chromatographic analytical methods: The illustrative case of androgenic steroids

This is a pre print version of the following article:

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1763238> since 2020-11-27T17:44:14Z

Published version:

DOI:10.1016/j.talanta.2020.120867

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

Manuscript Number:

Title: Effective validation of chromatographic analytical methods: the illustrative case of androgenic steroids

Article Type: Full Length Article

Section/Category: MASS SPECTROMETRY

Keywords: Validation; calibration; chromatography - mass spectrometry; androgens

Corresponding Author: Dr. Eleonora Amante,

Corresponding Author's Institution: Università degli Studi di Torino

First Author: Eugenio Alladio, PhD

Order of Authors: Eugenio Alladio, PhD; Eleonora Amante; Cristina Bozzolino; Fabrizio Seganti; Alberto Salomone; Marco Vincenti; Brigitte Desharnais, PhD

Manuscript Region of Origin: ITALY

Suggested Reviewers: Riccardo Leardi
Università di Genova - DIFAR
leardi@difar.unige.it

Mario Thevis
Centre for Preventive Doping Research (ZePräDo)
thevis@dshs-koeln.de

Francisco Raposo
Spanish National Research Council | CSIC · Industrial Processes & Environment Unit-Instituto de la Grasa
fraposo@cica.es

Pedro Araujo
National Institute of Nutrition and Seafood Research, PO
pedro.araujo@nifes.no



Università degli Studi di Torino

Dipartimento di Chimica

Via P. Giuria, 7 10125 Torino Italy



Eleonora Amante, MSc

phone: +39 3467891901

e-mail: eleonora.amante@unito.it

Torino, October 21th, 2019

Editor, **Analytica Chimica Acta**

Dear Editor,

This letter accompanies submission to *Analytica Chimica Acta* of a paper entitled “**Effective validation of chromatographic analytical methods: the illustrative case of androgenic steroids**” and of a parallel MethodsX paper entitled “**Experimental and statistical protocol for the effective validation of chromatographic analytical methods**”.

The authors are:

Eugenio Alladio, Eleonora Amante, Cristina Bozzolino, Fabrizio Seganti, Alberto Salomone, Marco Vincenti and Brigitte Desharnais.

Please address all correspondence to:

Eleonora Amante

Dipartimento di Chimica, Università degli Studi di Torino, Via Pietro Giuria 7, 10125, Torino (Italy)

Phone +39-3467891901

e-mail: eleonora.amante@unito.it

The validation of analytical methods is of crucial importance in several fields of application. The expounding case of a gas chromatographic-mass spectrometric method for the urinary endogenous steroid profiling is presented to illustrate a validation strategy that combines rigorous estimation of validation parameters with highly efficient use of the collected data. This work was inspired by two papers from Desharnais et al. published on the *Journal of Analytical Toxicology* (doi 10.1093/jat/bkx001 and 10.1093/jat/bkx002), which proposed a routine for the evaluation of calibration models. In practise, the validation protocol we describe requires three replicates of the calibration curve performed in three different days, for a total of nine replicates and 54 experiments. Such an operating scheme allows to evaluate several validation parameters using the same set of experiments. Among them, the calibration model is meticulously defined for each analyte, using several statistical tests for heteroscedasticity and linearity. With the same procedure, intra- and inter-day accuracy and precision are calculated.

This work provides an in-depth discussion of the results obtained with different statistical tools, using as representative example the case of a multi-targeted GC-MS method for the detection of androgen steroids in urine. All the equations and reported and described in the MethodsX parallel paper.

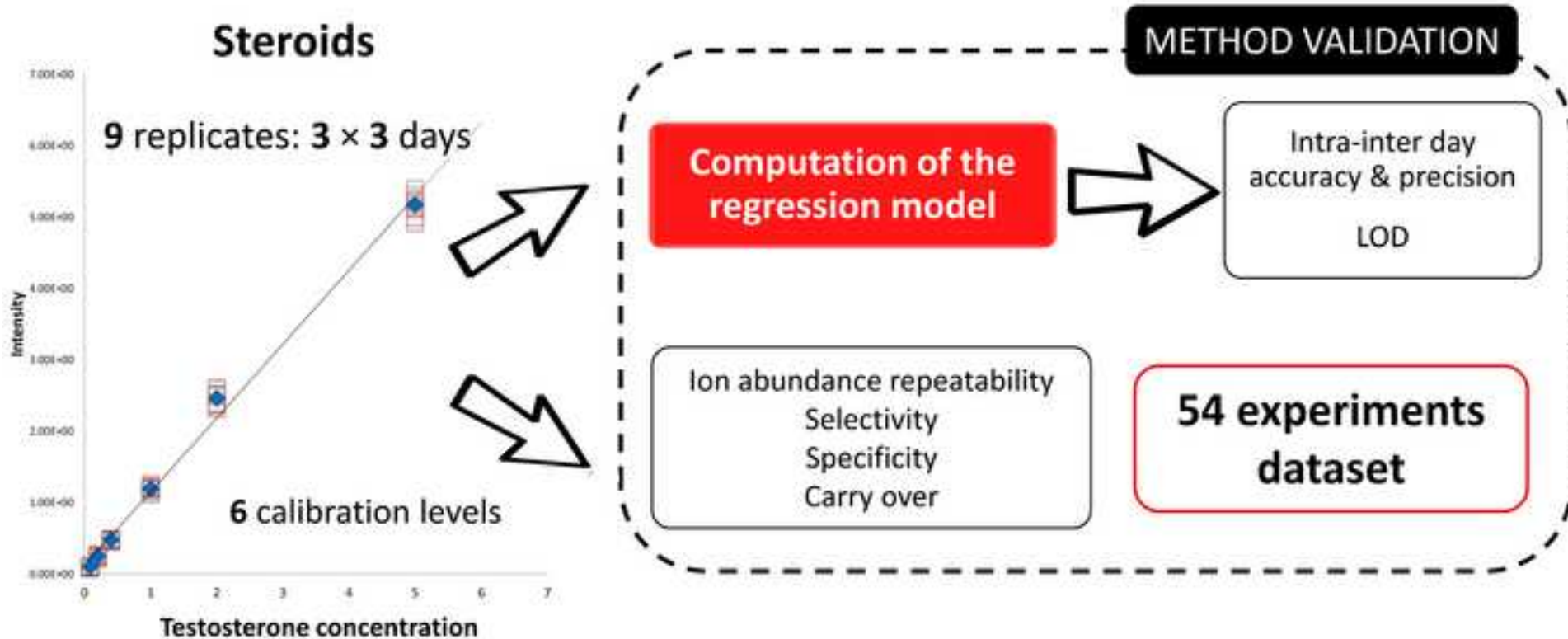
Novelty statement: This work is new and original and is not under consideration elsewhere. In the scientific literature, analytical method validations are frequently reported without clear relationship between the objectives of validation and the strategies of data collection and interpretation. Moreover, wrong statistical tools and unjustified assumptions are repeatedly used. We believe that our study represents an important tool of reflection for analytical chemists that can significantly contribute to standardize and improve the reliability of the validation process in the field of chromatography hyphenated with mass spectrometry.

Thank you for considering the paper for *Analytica Chimica Acta*.

Best regards.

Yours faithfully,

Eleonora Amante



Highlights

- The case study of a multitargeted method for urinary steroids is reported;
- An efficient and comprehensive validation strategy is proposed;
- From nine replicates of calibration data-points most validation parameters are calculated;
- Appropriate statistical tests are used and discussed.

Effective validation of chromatographic analytical methods: the illustrative case of androgenic steroids.

E. Alladio^{a,b*}, E. Amante^{a,b*#}, C. Bozzolino^{a,b}, F. Seganti^b, A. Salomone^{a,b}, M. Vincenti^{a,b}, B. Desharnais^c.

^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, Montréal, Québec, Canada

* The authors equally contributed to this work.

Corresponding Author:

Eleonora Amante,

Address: Dipartimento di Chimica, Università degli Studi di Torino, Via Pietro Giuria 7, 10125, Torino (Italy)

Telephone: +390116705264

Fax number: +390116705242

e-mail: eleonora.amante@unito.it

Keywords: Validation, calibration, chromatography – mass spectrometry, androgens.

Abstract

The increasing need to develop quantitative chromatographic methods with upgradable multi-targeted approach, allowing flexible and reliable application on large daily workload, makes the implementation of an efficient strategy of method's validation and maintenance crucial for the quality assurance policy. The expounding case of a gas chromatographic-mass spectrometric method for the urinary endogenous steroid profiling is presented to illustrate a validation strategy that combines rigorous estimation of validation parameters with highly efficient use of the collected data. The analysis of blank urine samples fortified at six concentration levels with 18 targeted steroids was replicated nine times in three working sessions along twelve days. This dataset of 54 analysis formed the groundwork on which the statistical evaluation of several validation parameters was founded, including calibration, intra- and inter-day accuracy and precision, limit of detection (LOD), limit of quantification, ion abundance repeatability, selectivity, specificity, and carry-over. The preliminary comparison of the response variances at different concentration levels provided the evaluation for heteroscedasticity. Then, the most appropriate calibration model was determined for each steroid, in terms of order (linear vs. quadratic) and weighting, allowing to complete their quantitation in each solution. Intra- and inter-day accuracy and precision were calculated therefrom. LOD values were computed with the Hubaux-Vos method from the weighted linear segment of the calibration curves. Only the assessment of recovery and matrix effect required the execution of further independent experiments. The case study demonstrated that the application of adequate statistical testing typically produced non-homogeneous models of calibration curves, mostly arising from heteroscedastic and quadratic distribution of datasets, unlike what is reported in overly simplified approaches. The misleading information obtained from the regression coefficient R^2 to evaluate linearity was evidenced. The strong dependence of calculated LOD and accuracy from the selected calibration parameters was highlighted, making the implementation of an adequate calibration maintenance policy highly advisable.

1. Introduction

The whole process of an analytical method validation has found various definitions. For example, the American Food and Drug Administration (FDA) defines it as “the process of demonstrating that an analytical procedure is suitable for its intended purpose” [1]. In practice, several organizations and scientists have tried to standardize the validation procedure, according to the purpose of the analysis (*e.g.*, qualitative, quantitative) and the application field, recommending specific parameters to be evaluated and tests to be performed [1–3]. Features of importance will differ depending on the particular application field (*e.g.* bioanalysis [4–6]) or instrumental technique used (*e.g.* chromatography [1,4]).

For the validation of quantitative methods, a feature of utmost importance is the calibration, which is the process that transforms the raw data obtained from the analytical instruments into useful concentration units by means of the statistical technique of regression [7]. Building an appropriate regression model requires the analysis of a series of standard samples within a defined range of concentrations, and the subsequent study of the mathematical relationships occurring between these concentrations and the corresponding analytical responses [7,8]. Consequently, the quality of the quantitative data that a novel analytical method will provide is highly dependent on the quality of the calibration model used [9]. Although most instrumental systems should theoretically exhibit analytical signals directly proportional to concentrations, hence generating linear calibration curves, in reality some interfering physical and chemical phenomena may result in a deviation from the expected linear trend [9,10] and/or heterogeneous distribution of data-point at different concentration levels [11]. To account for the latter problems, several publications and official documents have proposed the use of various statistical tests to compare linear and quadratic calibration curves and weighted least squares (WLS) regression strategies to correct heteroscedastic distributions [9,10,12–19]. The International Union of Pure and Applied Chemistry (IUPAC) guideline recommends to choose the model (*e.g.* ordinary linear regression, second order calibration

68 function or weighted regression) which provides the lowest measurement uncertainty [15].
1
269 Baumann and Watzig [14] developed a stepwise approach aimed to find the best weighted
3
470 calibration curve. In 2005, Singtoroj et al. [12] developed a systematic method to calculate and
5
6
71 compare regression models during pre-validation and validation of bioanalytical methods. Their
8
972 approach considered both quadratic and linear fitting, forcing through origin, transformation of data
10
11
1273 strategies (*e.g.* log-log, Box-Cox) and applied weighting (1, 1/x or 1/x²). More recently, Desharnais
13
1474 et al. [10,20] published an R routine devoted to the automatic testing and selection of the best
15
16
1775 calibration model, including order (linear or quadratic) and weighting (1, 1/x or 1/x²). In 2016,
18
1976 Raposo [21] reviewed the validation guidelines of several organizations and compared the tests
20
21
2277 recommended to evaluate fitting and linearity of the calibration curves, concluding that several
23
2478 different ways were valid to assess the linearity of calibration curves. The Scientific Working
25
2679 Group for Forensic Toxicology (SWGTOX), Irish National Accreditation Board (INAB), Joint
27
28
2980 Research Centre (JRC) and IUPAC all recommended executing several statistical tests within their
30
31
32 validation protocols.

33
34
3582 Another fundamental determination included in all validation procedures is the limit of detection
36
3783 (LOD), which is calculated either by examining the variance of the residual signal at zero
38
39
4084 concentration, using a large number of independent blank samples, or – more practically with
41
4285 chromatographic methods – by estimating its value from the lowest levels of the calibration curve
43
44
4586 [22]. The regression-based Hubaux-Vos' algorithm is a widespread technique for this estimation
46
4787 [22,23], but one of the prerequisites to use this approach is that the residuals from the linear model
48
49
5088 are homoscedastic, i.e. have a uniform distribution along the whole calibration range [13,22,23]. In
51
5289 contrast, the occurrence of heteroscedasticity is commonly observed in routine analytical models
53
54
5590 that cover several orders-of-magnitude concentrations. To overcome this problem, the use of an
56
5791 unweighted model can be replaced by a weighted least squares (WLS) calibration model in the
58
5992 Hubaux-Vos' LOD calculation [13].
60
61
62
63
64
65

93 The calculation of several other validation parameters is recommended in all guidelines, including
1
24 trueness, precision, specificity, recovery, matrix effect, carry over, and others [1–4,8,15,18,22]. The
3
4
55 evaluation of each of these parameters generally requires repeated independent experiments,
6
76 making the whole validation process quite demanding.

9
10
1197 In our daily work with biological matrices, LC-MS/MS and GC-MS-based methods are
12
1398 continuously developed and/or updated to support the ongoing evolution of clinical and
14
15
1699 toxicological requirements. To reduce the number of experiments needed to achieve a
17
1800 comprehensive validation of our analytical methods, we studied an integrated approach, resulting in
19
20
2101 the development of an efficient and rigorous validation protocol. This integrates the Desharnais’
22
2302 routine R procedure of calibration [10] into an inclusive strategy to estimate further validation
24
25
26103 parameters: intra- and inter-day accuracy and precision, LOD, limit of quantification (LOQ), ion
27
2804 abundance repeatability, selectivity, specificity and carry-over. At present, this protocol is routinely
29
30
3105 used in our laboratory for the validation of both GC-MS and UHPLC-MS/MS bioanalytical
32
3306 methods devoted to the determination of endogenous metabolites and xenobiotics [24–26]. In the
34
3507 present study, the whole validation strategy is presented, using as a case study the analytical method
36
37
3808 developed for the endogenous androgenic steroids.

41 42 43 44 4510 **2. Material and Methods**

46 47 48 49 50 5112 **2.1 Chemicals and reagents**

52
53
5413 All steroid standards were purchased as pure powders from Steraloids (Newport, RI, USA).
55
5614 Methanol, methyl tert-butyl ether (TBME), ethyl acetate, 17 α -methyltestosterone, dithioerythritol,
57
58
5915 N-Methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA) and synthetic urine were provided by
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

116 apparatus (Millipore, Bedford, MA, USA). Solid-phase extraction (SPE) C-18 endcapped cartridges
1
117 were from UCT Technologies (Bristol, PA, USA).
2
3
4
5

118 Standards solutions were prepared in methanol at the concentration of 1 mg/mL. Then, two working
7
119 solution mixtures were prepared by dilution (MIX I = 3 µg/mL, MIX II = 100 µg/mL, internal
9
120 standard solutions = 10 µg/mL). Two internal standards were used: testosterone-D₃ for the
12
121 quantification of mix I; 17α-methyltestosterone for mix II (Table 1).
14
15

1122 **2.2 Samples pre-treatment**

18
19

123 The investigated samples were either synthetic or negativized real urine specimen, depending on the
22
124 specific experiment. For the evaluation of repeatability of retention times and ion abundance ratios
24
125 authentic urine samples were negativized by extracting its steroid content by Solid-Phase Extraction
26
126 (SPE) using C-18 endcapped cartridges. The absence of analytes at concentrations above the limits
29
127 of detection was verified by the analysis of a non-spiked sample. Both authentic and synthetic urine
31
128 samples were treated identically. 6 mL urine aliquots were fortified with testosterone-D₃ and 17α-
34
129 methyltestosterone at the final concentration of 25 ng/mL and 125 ng/mL, respectively. The pH was
36
130 then adjusted to a value between 6.8 and 7.4 by adding 2 mL phosphate buffer 0.1 M and drop(s) of
38
131 NaOH 1 M, if necessary. A volume of β-glucuronidase solution corresponding to 83 units was
41
132 added and then the mixture was incubated at 58 °C for 1 hour. After cooling at room temperature, 2
43
133 mL carbonate buffer 0.1 M was added to the aqueous solution, together with drop(s) of NaOH 1 M,
46
134 until the final pH = 9 was reached. Then, liquid-liquid extraction (LLE) was performed with 10 mL
48
135 of TBME; the samples were shaken in a multi-mixer for 10 minutes, centrifuged at 6.24 x g for 5
51
136 minutes and the organic supernatant was transferred into a glass tube. The extracts were
53
137 subsequently dried under a nitrogen flow at 70 °C. After addition of 50 µL derivatizing solution
56
138 (MSTFA/NH₄I/dithioerythritol – 1,000:2:4 v/w/w), the reaction was allowed to proceed at 70 °C for
58
139 30 minutes. The resulting solutions were transferred into conical vials and a 1 µL aliquot was
60
61
62
63
64
65

140 injected by autosampler into the GC-MS working in the splitless mode. Further instrumental details
1
141 are available in a previously published version of this method [25] and in the MethodsX article
3
142 accompanying this work [27]. Mix I and II had distinct calibration ranges (Table 1), selected on the
6
143 basis of the expected physiological concentrations, as reported in literature [25,28].
8
9

10
11

12

13 2.3 Validation protocol

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

159 2.3.1 Calibration

160 A stepwise standard approach [10] was applied to calculate the best calibration models, as
54
161 schematized in Figure 1. Initially, the heteroscedasticity of data points was tested by comparing the
55
162 variance of the area ratios at the lowest and highest calibration level, using an F-Test integrated in
56
163 the R routine [20,29]. Also the Levene test [30] was executed (in the version modified by Brown
61

164 and Forsythe), in order to confirm the F-Test results with a procedure robust to non-normal
1
165 distributions, operating on all the calibration levels. If the variance increased with concentration, i.e.
2
166 the system was heteroscedastic, a weighted model was adopted, using a $1/x$ weighting factor when
3
167 the variance increased proportionally to the concentration, or a $1/x^2$ weighting when a quadratic
4
168 increase of the variance was observed. Case by case, the weighting generating the smallest variance
5
169 of weighted normalized variances was selected.
6
7
8
9
10
11
12
13
14
15

16 In the second step, the order of the calibration model was established by comparing the captured
17
18 variances of quadratic and linear (weighted) models by a partial F-test, which differs from Mandel's
19
20 test because it compares the sum of squares of the regression (not of the residuals) to the mean
21
22 squares of residuals [27]. If the quadratic calibration model significantly improved the captured
23
24 variance of the data in comparison with the linear model, the former was accepted and the algorithm
25
26 computed it. Lastly, the analysis of variance lack of fit (ANOVA-LoF) test was performed, to verify
27
28 the goodness of the calculated calibration model.
29
30
31
32
33

34 2.3.2 LOD

35
36
37 The limit of detection (LOD) was estimated by means of the Hubaux-Vos' algorithm [23], which is
38
39 an implementation of Currie's method to calculate the LOD [31]. This technique estimates the
40
41 concentration associated with the confidence interval for the signal of a blank sample (i.e. the
42
43 intercept). This is a robust means to estimate the LOD, since it relies on the entire data set to do so.
44
45
46 Although the original Hubaux-Vos technique applies to homoscedastic data, Sanchez [13] has
47
48 demonstrated that it can be used for heteroscedastic data if properly weighted parameters (standard
49
50 error of the regression, slope and intercept, etc.) are used. Moreover, to respect Hubaux-Vos
51
52 linearity assumption, when a quadratic calibration trend was recorded, the highest concentration(s)
53
54 (typically the last level) were discarded from the calculation until a linear dependence from
55
56 concentration was observed, as confirmed by the same partial F-test previously employed for the
57
58
59
60
61
62
63
64
65

188 definition of the calibration model. The calculated LOD values were experimentally tested by
1
189 spiking the blank matrix (synthetic urine) with the targeted analytes at the approximate LOD
3
190 concentration and verifying that the signal-to-noise ratio (S/N) was higher than 3.
4
5
6
7

191 **2.3.3 Accuracy**

8
9

10
11
192 In most validation protocols reported in the literature, accuracy and trueness [32] are calculated by
13
193 executing at least five determinations per concentration at a minimum of three concentration levels,
14
15
194 and a deviation below 15% from the expected value [33] is recommended as an acceptance
16
17
195 criterion. In contrast, our routine relies on an independent back-calculation of each data-point using
18
19
20
21
296 only the calibration curves that did not include the sequence containing the specific data-point
22
23
247 considered. The operating scheme is summarized in Figure 2A. For the calculation of intra-day
24
25
26
198 accuracy, each sequence of six-level analysis was imagined to provide a separate calibration line,
27
28
299 theoretically yielding three calibrations per each day. Then, (i) two sequences were used to compute
29
30
31
300 the calibration model using the R routine, as described above; (ii) this model (which might be
32
33
34
301 weighted or unweighted, linear or quadratic, depending on the former conclusions) was used to
35
36
302 back-calculate the data-points of the third sequence; (iii) all concentration results from the back-
37
38
303 calculation were averaged, and the overall bias was calculated.
39
40
41

42
304 The inter-day accuracy was computed following a similar operating scheme (Figure 2B), where the
43
44
45
46
47
306 back-calculations of the data-points from a specific day were performed on the calibration model
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165

166 **2.3.4 Precision**

211 Precision was expressed by the coefficient of variation (%CV or %RSD) of calculated
1
212 concentrations from repeated analysis of homogeneous urine aliquots [33–35] spiked to provide the
3
4
213 calibration curves. The intra-assay precision was calculated independently for the three days of
5
6
214 analysis using the three replicates obtained in each day. A calibration model was computed for each
7
8
9
215 day of validation and it was used to back calculate the three experimental replicates performed the
10
11
216 same day. Then, the %CV was calculated [27].
12
13
14

15
217 The inter-assay (or intermediate) precision, which applies to within laboratory variations [36], was
16
17
218 computed by back calculating all the nine replicates using the comprehensive calibration curve (i.e.
18
19
219 the one built using all the performed experiments). Our protocol allowed to compare the
20
21
22
220 performance of the same analysis in different days and by different operators, involving two
23
24
25
221 operators working in alternate sequence.
26
27

28
222 Satisfactory results were expected to lie within $\pm 15\%$ for both intra and inter-assay precision.
29
30

31 32 33 **2.3.5 Selectivity and Specificity** 34 35

36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

The presence of potentially interfering substances in urine, including endogenous matrix components, metabolites, and decomposition products, was checked by examining the selected ion chromatograms around the expected retention times for all the analytes of interest. The presence of interfering peaks with $S/N > 3$ around the retention time of the analytes was examined for all samples in each experiment. Identification criteria for the analytes were established by checking the presence of all qualifying ions and their relative abundance at the expected retention time.

66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

2.3.6 Repeatability of retention times and ion abundance ratios
Retention time repeatability was evaluated using the calibration standards and 30 blank (negativized) real urine samples fortified with the target analytes at different concentration levels within the dynamic range of the analytical methodology. The latter were prepared from 6 mL real

234 urine aliquots by extracting the steroid fraction by C-18 SPE and subsequently spiking the eluates
1
235 with the target analytes. Deviations below 1% from calibrators and controls are usually considered
3
4
236 acceptable. Ion abundance ratio (quantifying to qualifying ion) repeatability was evaluated for each
5
6
237 target analyte at all calibration levels, with acceptance limit of $\pm 20\%$ with reference to the control.
8
9

238 **2.3.7 Matrix effect**

10
11
12
13
14
239 The matrix effect was estimated at three concentration levels (e.g. low, medium and high
15
16
240 concentration, within the linear range of the method) by comparing the experimental results
18
19
241 obtained from synthetic blank urine samples and blank deionized water samples, both spiked after
20
21
242 the extraction step. The matrix effect for each target analyte was expressed as the percentage ratio
23
243 between the two measured concentrations.
25
26

244 **2.3.8 Extraction recovery**

27
28
29
30
31
245 The extraction recovery was determined by comparing the experimental results obtained from
32
33
246 synthetic urine samples respectively spiked before and after the extraction step. It was expressed as
35
36
247 the percentage ratio between the two quantified concentrations and estimated at the first, third and
37
38
248 last calibration levels, in triplicate, for a total of 18 samples. In practice, only the experiments
40
41
249 involving the spiking after the extraction step were added to the sequences performed to build the
42
43
250 calibration curves.
44
45
46

251 **2.3.9 Carry-over effect**

47
48
49
50
252 The carry-over was evaluated by injecting distilled water extracts after the highest point of each
52
53
253 calibration curve. If the signal-to-noise ratio was lower than 3 for each selected ion, the carry-over
54
55
254 effect was considered negligible.
57
58

255 **3. Results and Discussion**

256 **3.1 Validation protocol**

1
2
3
257 To exemplify the step-by-step validation protocol implemented within our quality assurance policy,
4
5
258 the analytical method based on gas chromatography-mass spectrometry (GC-MS) for the detection
7
8
259 of 18 endogenous anabolic androgenic steroids (EAAS) in the urine of male individuals is used in
9
10
260 this paper. This analytical procedure was developed for diagnostic purposes, to screen individuals
12
13
261 with suspected prostate cancer [25]. All monitored steroids are reported in Table 1, with their CAS
14
15
262 numbers, retention times and internal standard. Validation results obtained for all analytes are
16
17
263 reported in Tables 2 to 6 and in the Supplementary Material. An in-depth discussion of the
19
20
264 procedure is conducted in the following paragraphs for three steroids representing different and
21
22
265 emblematic casework conditions:

- 26
266 - A target analyte (testosterone (T)) quantified using the corresponding isotopically-labelled
27
28
267 homologue (testosterone-d3 (T-d3));
30
- 31
268 - A target analyte (4,6-androstadien-3,17-dione (6-D)) quantified using an isotopically-
32
33
269 labelled compound with different structure (testosterone-d3), but belonging to the same
35
36
270 chemical class (*viz.* C-19 steroids);
37
- 38
271 - A target analyte (androsterone (Andro)) quantified in a higher and wider calibration range
40
41
272 using an exogenous compound (17 α -methyltestosterone (17 α -methyl-T)), not isotopically-
42
43
273 labelled.
44

45
46
47
274 These three cases were selected as representative of experimental conditions commonly found in
48
49
275 most analytical laboratories. The first condition is ideal, but it is not always feasible, for example
51
52
276 due to the unavailability of isotopically-labelled standards for all members of a large series of target
53
54
277 analytes. Conversely, the second and third conditions are easier to fulfil and are largely adopted by
56
278 laboratories.
58

59
60
279 **3.2 Linear Dynamic Range and Calibration**
62
63
64
65

280 The linear instrumental response was evaluated within the concentration range $C_S = 2\text{-}125$ ng/mL
1
281 for mix I, and $C_S = 100\text{-}2250$ ng/mL for mix II. The calibration curves of T, 6-D and Andro are
2
3
4
282 reported in Figure 3, the values on the X-and Y-axis being C_S/C_{IS} and A_S/A_{IS} , respectively (where
5
6
283 C_{IS} is the fixed concentration of the internal standard -considered adimensional-, A_S and A_{IS} are the
7
8
284 area of the steroid and internal standard chromatographic peaks, respectively). The slope of the
9
10
11
1285 curves represents the response factor for each analyte and it is close to unity for T, as expected.
13
14

15
16
286 Although T and 6-D were spiked using the same working solution, the data-point variability
17
1887 exhibited by 6-D is considerably higher than what is observed for T. The same phenomenon occurs
19
20
288 for other analytes in MIX I, with the exception of E (T-epimer). This reflects the parallel signals
21
22
289 fluctuation of T and E with T-d3, which does not occur for the other analytes. It appears that the
23
24
25
290 overall variability of experimental results increases as the difference between the analyte and
26
27
291 internal standard structure becomes more imposing. Following this trend, data collected for the
28
29
30
292 calibration of Andro is more scattered than for 6-D, since Andro and 17α -methyl-T display a larger
31
32
293 structural difference than 6-D and T-d3.
33
34
35

36
37
38
394 All data-point distributions were found to be heteroscedastic, as indicated by congruous results from
39
40
41
42
43
44
45
46
47
48
49
50
500 is apparently more pronounced for Andro, as the response variance becomes quite pronounced at
51
52
53
501 the high concentration levels. This result can be attributed to the combined effect of structural
54
55
56
57
58
59
60
604 difference with the internal standard, as discussed above, and the higher concentration range
61
62
63
64
65 explored with respect to it. On the other hand, the significance of the heteroscedasticity F-test is
extremely high for T, despite the limited spread of data-points observed at any concentration.

305 Indeed, for T and 6-D, the distributions at the lower concentration levels (from 2 to 25 ng/mL)
1
306 appear to be almost homoscedastic, but a wider spread of the replicates is observed at the higher
3
4
307 concentration levels (50 and 125 ng/mL). The described trends have been highlighted in previous
6
308 studies [13,18].
8
9

10
309 The order of the calibration model was subsequently chosen. For most analytes (including T and 6-
12
310 D) a quadratic model proved superior, while the introduction of the second order term turned out
14
311 negligible for Andro and few of other target analytes, making the linear fitting more advisable for
16
17
312 them. All the equations for the final calibration models are reported in Table 2. These models
19
20
313 showed the best performance under the partial F-test for the quadraticity.
21
22
23

24
314 The ANOVA-LoF was computed for all the 18 analytes using the complete set of validation. This
25
26
315 test was intended to verify the fit of data-points with the final calibration model, irrespective of its
27
28
316 order and weighting. For 10 of them, the calibration model was rejected, despite the good accuracy
30
31
317 and precision performances (see Table 6 and Supplementary Table 1). This is in line with the
32
33
318 findings from Desharnais *et al.* [20], who concluded that the excessive sensitivity of this test to the
35
36
319 chosen experimental design (in particular the number of replicates and of calibration levels) limits
37
38
320 its practical applicability. Hence, we can consider not to use the outcome of this testing as a strict
40
41
321 acceptance criterion within the validation protocol.
42
43
44

45
322 Even if this is widely done in method validation procedures and in the literature, Table 2
46
47
323 demonstrates that the adoption of identical calibration models (weighting factor, order) for large
48
49
324 sets of target analytes does not provide the best fit for the data. Quite often, the software for data
51
52
325 analysis provided with most instrumentation encourages the use of homogeneous models by asking
53
54
326 their definition prior to testing. Additionally, insufficient statistical testing frequently characterizes
55
56
327 the method validations reported in the literature.
58
59
60
61
62
63
64
65

328 Comparison between Figure 3 and Table 2 highlights the misleading role played by the squared
1
329 correlation coefficient (R^2 , also called global goodness of fit) [37,38] in the common practice of
3
4
330 using it as a test for linearity. Although R^2 values higher than 0.99 were calculated for the linear
5
6
331 calibration of T and 6-D (Figure 3) and could have been interpreted as a confirmation of the validity
7
8
332 of the linear model, both visual inspection and Partial F-test test outcomes (Table 2) clearly indicate
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

334 After the choice of the calibration model was completed, back-calculation of the concentration for
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

3.3 LOD and LOQ

41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

352 When quadraticity is noted in the data, several calibration options are available: (i) use the quadratic
1
353 calibration model to quantify the real samples; (ii) split the calibration into two ranges both fitted
3
4
354 with a linear model; (iii) reduce the calibration range to a narrower linear interval. Whereas the
6
355 second approach might be advisable whenever the calibration range covers two or more orders of
8
356 magnitude, the latter strategy is a relevant choice to calculate the LOD. Figure 4 shows linear
10
11
1357 regressions for T and Andro with and without the sixth and highest calibration level. Obviously, the
13
14
358 intercept of the green curve is closer to the origin of the y-axis, which is closer to the expected
15
16
1359 outcome and results in lower LOD values and reduced uncertainty in their estimation.

18
19
20
21360 To illustrate this concept, LOD values were calculated using linear models and 6, 5, and 4
22
2361 calibration levels, respectively (Table 4). For T and 6-D, the LOD values are 2- to 4-times lower
24
25
26362 when computed using the five-points calibration models instead of the six-points, while the LOD
26
27
28363 for Andro remained virtually unaffected, as expected. This is in agreement with previous studies,
29
30
31364 which demonstrated that LODs are overestimated whenever deviation from homoscedasticity and
32
3365 linearity are ignored [9,12]. Similar trends were observed for the remaining steroids: LOD values
34
35
36366 generally tend to decrease when the concentration levels used within the Hubaux-Vos method are
36
37
38367 reduced to the lowest calibration levels. On the other hand, this increases the risk of LOD
39
4068 underestimation and makes the experimental verification with blank samples spiked at the LOD
41
42
43369 level highly advisable. The results of this verification are reported in Table 4, where the most
44
4570 reliable calculated values are displayed in bold. This dependable LOD value was chosen after
46
47
48371 careful inspection of the data point distribution and experimental confirmation.

49
50
5172 In the present method, targeted analytes were endogenous steroid, whose actual concentration in
52
53
5473 real samples may largely exceed the LOD values. While LOQ values could theoretically be defined
55
5674 as two or three times the LOD, we preferred to assess them as the minimal concentrations allowing
57
58
5975 to guarantee quantitative determinations with acceptable accuracy, within the range of physiological
60
6176 values. This makes LOQ and accuracy strictly conjoined concepts, as discussed below.

377
1
2
378
3
4
5
379
7
8
380
9
10
381
12
13
382
14
15
383
17
18
19
384
20
21
22
385
23
24
386
25
26
387
28
29
388
30
31
389
32
33
390
35
36
391
37
38
392
40
41
393
42
43
394
44
45
395
46
47
48
49
396
50
51
397
52
53
54
398
55
56
399
57
58
400
59
60
61
62
63
64
65

3.4 Accuracy and precision

An interesting feature of the proposed method validation procedure is that the large data set collected throughout several days for calibration purposes can be exploited to calculate intra and inter-day accuracy and precision at six concentration levels (instead of the common low, intermediate, and high levels) without requiring any further experimental work. These results for the studied method are reported in Tables 5 and 6.

The first highlight obtained from Table 5 is that none of the 216 accuracy results exceeds ± 25 , whereas 18 (8.3 %) exceed ± 20 , 21 (9.7 %) are comprised between ± 15 and ± 20 , and the remaining 177 (82%) are within ± 15 . Only two substances (4-androsten-3,17-dione and formestane) display intra-day and inter-day accuracy which are not satisfactory at several concentrations. Beside these analytes, the overall accuracy of the method is acceptable and even fully satisfactory for most of the analytes, yet not perfect, as is expected for a wide set of authentic results. In particular, 7 out of the 18 accuracy results exceeding ± 20 are found in the highest concentration level, while the others are randomly scattered throughout the other concentration levels. Again, this is to be expected since the calibration uncertainty is maximal at the extremes of the concentration range. Notably, the six steroids included in the steroidal module of the WADA Athlete Biological Passport (T, Andro, E, Etio, 5α -adiol, 5β -adiol), consistently provided accuracy data largely below 10, except 5α -adiol and 5β -adiol at the highest concentration level.

Another interesting observation is that inter-day accuracy frequently proved higher than intra-day accuracy, except for 7α -hydroxytestosterone. Together with the absence of any day-specific clustering of data-points in the nine-replicated analysis, this leads to the conclusion that under stable instrumental (GC-MS) conditions, it is preferable to use averaged calibration curves built from data collected on different days rather than changing the calibration curve daily with a new set of

standards. Considering that the first and third sets of experiments were separated by twelve days, it is safe to assume that analytical data from standard solutions can be collected over a two weeks interval to build robust calibration curves from a large number of replicates. Possibly, these conclusions may not hold for LC-MS/MS methods, where large inter-day variation of the signal is commonly observed.

The tests used to verify intra-day and inter-day precision overall mirrors the conclusions drawn from the accuracy testing. Table 6 displays the data for the steroids included in the WADA Athlete Biological Passport, plus 6-D, while the rest of the variation coefficients is provided in the Supplementary Material Table S1.

Similarly to accuracy data, the precision performance was highest for T (likely due to the use of the isotopically labelled analogue T-d3 as the internal standard for signal correction), followed by the group of steroids included in the WADA list for the Athlete Biological Passport. All these steroids are either metabolites or direct precursors of T and have closely related chemical structures, hence yields in the critical steps of sample processing (extraction, derivatization, *etc.*) are likely more closely correlated with T-d3, thus achieving better signal correction.

All together, relatively good performances were obtained in terms of inter-day accuracy and precision, suggesting that the analytical method is robust and the response of the GC-MS instrument is stable within a two weeks' time frame. On a daily basis, a restricted number of standards can be analysed to evaluate the stability of the instrument response in the routine activity before starting the analytical session with real samples. The obtained results may serve the purpose of quality control for the current session, while new calibration data obtained afterwards could possibly be added to the ongoing calibration, to replace the oldest data. For example, updated calibration curves could be obtained in triplicate each week, and the corresponding validation parameters recalculated automatically using a worksheet or a routine such as the ones described elsewhere [10,27]. The

425 quality control results obtained using the existing calibration confirms that correct operating
1
426 conditions are maintained or, if the confidence limits are exceeded, that careful revision of the
2
3
4
427 analytical steps and recalibration are needed.
5
6
7

8
428 In general, several method maintenance strategies are compatible with the validation scheme
9
10
429 presented herein, depending on the specific method features and validation results. For example, the
11
12
430 ongoing control of intra-day accuracy and precision may require a higher number of repetitions than
13
14
431 the three sets of three repetitions proposed above. On the other hand, meticulous verification of
15
16
432 intra-day variability may prove excessive if inter-day parameters turn out to be wholly under
17
18
19
20
433 control.
21
22
23

24 3.5 Other validation parameters 25

26
27
28
435 Retention time repeatability proved satisfactory, as ascertained by both the calibration analyses and
29
30
436 those made on fortified negativized samples (see Materials and Methods). No significant deviation
31
32
437 from the expected retention time was observed. Repeatability of ion abundance ratios, tested for all
33
34
438 the calibration levels, provided results within the limits of acceptance ($\pm 20\%$).
35
36
37

38
39
40
41
440 Likewise, the analytical method proved selective and specific for all the targeted compounds, since
42
43
441 no interfering peaks appeared in any real urine sample around the retention time of the analytes with
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

447 Lastly, no evidence of carry-over was detected, since the blank samples injected after the higher
1
448 level of calibration showed no signal (*viz.* $S/N < 3$) for all the analytes.
3

4
449

6

450 **4. Conclusions**

8

9

10

451 The case study presented hereby investigated the complexity of developing a rigorous validation
12

13

452 procedure for chromatographic methods devoted to multi-analyte targeted quantitative
14

15

453 determinations of (endogenous) analytes. This complexity arises from the lack of homogeneity of
17

18

454 the data amongst the different analytes, the need to verify several performance properties with
19

20

455 serious statistical testing along long periods of potential application, and finally the need to ease the
21

22

456 daily routine work when high throughput is required.
24

25

457 To meet these requirements, we proposed an operating protocol whose core consists in the
26

27

458 systematic replication of three calibration curves in three different days (for a total of nine
29

30

459 replicates) within a time lapse of 12 days. This protocol allowed robust evaluation of the calibration
31

32

460 curves and simultaneous assessment of other validation parameters, including LOD, LOQ, intra-
34

35

461 and inter-day precision and accuracy, ion abundance repeatability, selectivity, specificity and carry
36

37

462 over.
38

39

40

463 Along this case study, we demonstrated that the application of different statistical tests for linearity
42

464 and homoscedasticity typically produced non-homogeneous results across analytes and tests, which
44

45

465 have to be evaluated independently with care before drawing any conclusion. In general,
46

466 heteroscedastic and quadratic distribution of calibration data-sets were more frequent than linear
48

49

50

467 trends, except when very limited concentration ranges were considered. Even analytes with similar
51

52

468 chemical structures (*i.e.*, androgenic steroid) and similar concentration ranges may require different
54

55

469 calibration criteria and should be selected case by case, unlike what is reported in numerous
56

57

470 literature reports which appear overly simplified and optimistic. Moreover, the inadequacy of the
59

60

61

62

63

64

65

471 regression coefficient R^2 to evaluate the linearity was demonstrated once more: quadratic and
1
472 heteroscedastic data distribution proved compatible with high and misleading R^2 values.
3
4

473 The interdependence of LOD and LOQ values with calibration, accuracy, and precision parameters
5
6
7
474 has been clearly shown. If LOD values are calculated partly or completely from the calibration
8
9
10
475 regression equation, for example because real blank samples are lacking, then the number of
11
12
476 concentration levels used for the regression and the distribution of data-points may strongly
13
14
477 influence the outcome. An experimental confirmation of LOD and LOQ values and their
15
16
17
478 verification with accuracy and precision testing is always advisable.
18
19

20
479 Lastly, our data showed that the collection of multiple calibration results, their averaging, and their
21
22
480 continuous refreshing within a quality control process produced more accurate quantitation than the
23
24
25
481 use of a single calibration collected on a daily basis.
26
27
28
29
30
31

32 **References**

33
34

- 35 [1] U.S. Department of Health and Human Services, Food and Drug Administration (FDA),
36
37 Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and
38
39 Research (CBER), U.S. Department of Helth and Human Services, Guidance for Industry:
40
41 Analytical Procedures and Methods Validation for Drugs and Biologics (Draft Guidance),
42
43 (2014) 1–14.
44
45 <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidance>
46
47 [s/UCM386366.pdf](http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidance).
48
49
50
51
52 [2] I. Taverniers, M. De Loose, E. Van Bockstaele, Trends in quality in the analytical laboratory.
53
54 II. Analytical method validation and quality assurance, TrAC - Trends Anal. Chem. 23
55
56 (2004) 535–552. doi:10.1016/j.trac.2004.04.001.
57
58
59
60 [3] G. Shabir, Step-by-step analytical methods validation and protocol in the quality system
61
62
63
64
65

- 495 compliance industry, *J. Valid. Technol.* 10 (2004) 210–218. doi:10.1038/ncomms11248.
- 1
2
3
496 [4] O. González, M.E. Blanco, G. Iriarte, L. Bartolomé, M.I. Maguregui, R.M. Alonso,
4
5
6
7
8
9
10
11
12
13
1499 [5] M. Kaza, M. Karaźniewicz-Łada, K. Kosicka, A. Siemiątkowska, P.J. Rudzki, Bioanalytical
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- [6] D.E. Coleman, L.E. Vanatta, Lack-of-fit testing of ion chromatographic calibration curves with inexact replicates, in: *J. Chromatogr. A*, Elsevier, 1999: pp. 43–51. doi:10.1016/S0021-9673(99)00369-6.
- [7] L.E. Vanatta, D.E. Coleman, Calibration, uncertainty, and recovery in the chromatographic sciences, *J. Chromatogr. A*. 1158 (2007) 47–60. doi:10.1016/j.chroma.2007.02.040.
- [8] E. Rozet, A. Ceccato, C. Hubert, E. Ziemons, R. Oprean, S. Rudaz, B. Boulanger, P. Hubert, Analysis of recent pharmaceutical regulatory documents on analytical method validation, *J. Chromatogr. A*. 1158 (2007) 111–125. doi:10.1016/j.chroma.2007.03.111.
- [9] A.M. Almeida, M.M. Castel-Branco, A.C. Falcão, Linear regression for calibration lines revisited: Weighting schemes for bioanalytical methods, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 774 (2002) 215–222. doi:10.1016/S1570-0232(02)00244-1.
- [10] 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.
- [11] A. Sayago, A.G. Asuero, Fitting straight lines with replicated observations by linear regression: Part II. Testing for homogeneity of variances, *Crit. Rev. Anal. Chem.* 34 (2004)

- 519 133–146. doi:10.1080/10408340490888599.
1
2
520 [12] T. Singtoroj, J. Tarning, A. Annerberg, M. Ashton, Y. Bergqvist, N.J. White, N. Lindegardh,
4
521 N.P.J. Day, A new approach to evaluate regression models during validation of bioanalytical
6
522 assays, *J. Pharm. Biomed. Anal.* 41 (2006) 219–227. doi:10.1016/j.jpba.2005.11.006.
8
9
10
523 [13] J. Sanchez, Estimating Detection Limits in Chromatography from Calibration Data: Ordinary
11
12
524 Least Squares Regression vs. Weighted Least Squares, *Separations*. 5 (2018) 49.
14
15
525 doi:10.3390/separations5040049.
16
17
18
526 [14] K. Baumann, H. Wätzig, Appropriate calibration functions for capillary electrophoresis II.
19
20
527 Heteroscedasticity and its consequences, *J. Chromatogr. A*. 700 (1995) 9–20.
21
22
528 doi:10.1016/0021-9673(95)00128-A.
24
25
529 [15] S. Bratinova, B. Raffael, C. Simoneau, Guidelines for performance criteria and validation
27
28
530 procedures of analytical methods used in controls of food contact materials, 2009.
29
30
531 doi:10.2788/49046.
32
33
532 [16] C.P. Da Silva, E.S. Emídio, M.R.R. De Marchi, Method validation using weighted linear
35
36
533 regression models for quantification of UV filters in water samples, *Talanta*. 131 (2015) 221–
37
38
534 227. doi:10.1016/j.talanta.2014.07.041.
39
40
41
535 [17] S. Rudaz, M. Feinberg, From method validation to result assessment: Established facts and
42
43
536 pending questions, *TrAC - Trends Anal. Chem.* 105 (2018) 68–74.
45
46
537 doi:10.1016/j.trac.2018.04.013.
47
48
49
538 [18] F. Schiller, D. Jena, L. a Currie, International Union of Pure and Applied Chemistry.
50
51
539 Guideline for calibration in analytical chemistry- Part 1 . Fundamentals and single
52
53
540 component, *Pure Appl. Chem.* 70 (1998) 993–1014. doi:10.3390/rs8030220.
55
56
57
541 [19] J. Scott Long, L. Ervin, Correcting for Heteroscedasticity with Heteroscedasticity Consistent
58
59
542 Standard Errors in the Linear Regression Model: Small Sample Considerations, *Indiana*
60
61
62
63
64
65

- 543 Univ. Bloom. 47405 (1998) 1–33.
1
2
3
544 [20] B. Desharnais, F. Camirand-Lemyre, P. Mireault, C.D. Skinner, Procedure for the selection
4
545 and validation of a calibration model II-theoretical basis, *J. Anal. Toxicol.* 41 (2017) 269–
6
546 276. doi:10.1093/jat/bkx002.
7
8
9
10
11 547 [21] F. Raposo, Evaluation of analytical calibration based on least-squares linear regression for
12
13 548 instrumental techniques: A tutorial review, *TrAC - Trends Anal. Chem.* 77 (2016) 167–185.
14
15 549 doi:10.1016/j.trac.2015.12.006.
16
17
18
19 550 [22] E. Bernal, Limit of Detection and Limit of Quantification Determination in Gas
20
21 551 Chromatography, *Adv. Gas Chromatogr.* (2014). doi:10.5772/57341.
22
23
24 552 [23] A. Hubaux, G. Vos, Decision and Detection limits for linear Calibration Curves, *Anal.*
25
26 553 *Chem.* 42 (1970) 849–855. doi:10.1021/ac60290a013.
27
28
29 554 [24] A. Salomone, J.J. Palamar, R. Bigiarini, E. Gerace, D. Di Corcia, M. Vincenti, Detection of
30
31 555 Fentanyl Analogs and Synthetic Opioids in Real Hair Samples, *J. Anal. Toxicol.* (2018).
32
33 556 doi:10.1093/jat/bky093.
34
35
36
37 557 [25] E. Amante, E. Alladio, A. Salomone, M. Vincenti, F. Marini, G. Alleva, S. De Luca, F.
38
39 558 Porpiglia, Correlation between chronological and physiological age of males from their
40
41 559 multivariate urinary endogenous steroid profile and prostatic carcinoma-induced deviation,
42
43 560 *Steroids.* 139 (2018) 10–17. doi:10.1016/j.steroids.2018.09.007.
44
45
46
47 561 [26] E. Alladio, G. Biossa, F. Seganti, D. Di Corcia, A. Salomone, M. Vincenti, M.R.
48
49 562 Baumgartner, Systematic optimization of ethyl glucuronide extraction conditions from scalp
50
51 563 hair by design of experiments and its potential effect on cut-off values appraisal, *Drug Test.*
52
53 564 *Anal.* (2018). doi:10.1002/dta.2405.
54
55
56
57 565 [27] E. Alladio, E. Amante, C. Bozzolino, F. Seganti, A. Salomone, M. Vincenti, B. Desharnais,
58
59 566 Experimental and statistical protocol for the effective validation of chromatographic
60
61
62
63
64
65

- 567 analytical methods, MethodsX. Submitted.
1
- 2
568 [28] P. Van Renterghem, P. Van Eenoo, H. Geyer, W. Schänzer, F.T. Delbeke, Reference ranges
4
569 for urinary concentrations and ratios of endogenous steroids, which can be used as markers
6
570 for steroid misuse, in a Caucasian population of athletes, *Steroids*. 75 (2010) 154–163.
8
9
571 doi:10.1016/j.steroids.2009.11.008.
11
- 12
572 [29] B. Desharnais, F. Camirand-Lemyre, P. Mireault, C.D. Skinner, Procedure for the Selection
14
573 and Validation of a Calibration Model I—Description and Application, *J. Anal. Toxicol.* 41
16
574 (2017) 261–268. doi:10.1093/jat/bkx001.
19
- 20
575 [30] Levene, Robust tests for equality of variances, in: I. Olkin, H. Hotelling (Eds.), *Contrib. to*
22
576 *Probab. Stat. Essays Honor Harold Hotell.*, Stanford University Press, 1960: pp. 278–292.
24
- 25
577 [31] L.A. Currie, Detection and quantification limits: Origins and historical overview, *Anal.*
27
578 *Chim. Acta.* 391 (1999) 127–134. doi:10.1016/S0003-2670(99)00105-1.
29
- 30
579 [32] V.P. Shah, K.K. Midha, J.W.A. Findlay, H.M. Hill, J.D. Hulse, I.J. Mcgilveray, G. Mckay,
32
580 K.J. Miller, R.N. Patnaik, M.L. Powell, A. Tonelli, C.T. Viswanathan, *Bioanalytical Method*
35
581 *Validation—A Revisit with a Decade of Progress*, 17 (2000).
37
- 38
582 [33] M.R. Lakshmi HimaBindu, S. Angala Parameswari, C. Gopinath, A review on GC-MS and
40
583 method development and validation, *Int. J. Pharm. Qual. Assur.* 4 (2013) 42–51.
42
- 43
584 [34] S. V. Karmarkar, Validation of Ion Chromatographic Methods, *Appl. Ion Chromatogr.*
45
585 *Pharm. Biol. Prod.* (2012) 285–308. doi:10.1002/9781118147009.ch15.
48
- 49
586 [35] G. a Shabir, Validation of high-performance liquid chromatography methods for
51
587 pharmaceutical analysis, *J. Chromatogr. A.* 987 (2003) 57–66. doi:10.1016/S0021-
53
588 9673(02)01536-4.
55
- 56
589 [36] A.G. González, M.Á. Herrador, A.G. Asuero, Intra-laboratory assessment of method
58
59
600 accuracy (trueness and precision) by using validation standards, *Talanta*. 82 (2010) 1995–
61
62
63
64
65

591 1998. doi:10.1016/j.talanta.2010.07.071.

1

2

592 [37] Scientific Working Group for Forensic Toxicology (SWGTOX), Scientific working group

4

593 for forensic toxicology (SWGTOX) standard practices for method validation in forensic

6

594 toxicology, *J. Anal. Toxicol.* 37 (2013) 452–474. doi:10.1093/jat/bkt054.

7

8

595 [38] P. Araujo, Key aspects of analytical method validation and linearity evaluation, *J.*

10

596 *Chromatogr. B Anal. Technol. Biomed. Life Sci.* 877 (2009) 2224–2234.

12

597 doi:10.1016/j.jchromb.2008.09.030.

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

	Target analyte	CAS number	T _R (min)	Internal standard		
Mix I	5β-androstan-13,17-dione	1229-12-5	8.15	Testosterone-D ₃		
	5α-androstane-3α,17β-diol (5α-adiol)	1852-53-5	9.32	Testosterone-D ₃		
	5β-androstane-3α,17β-diol (5β-adiol)	1851-23-6	9.40	Testosterone-D ₃		
	dehydroepiandrosterone (DHEA)	53-43-0	10.00	Testosterone-D ₃		
	5-androsten-3,17-diol	512-17-5	10.24	Testosterone-D ₃		
	epitestosterone (E)	481-30-1	10.35	Testosterone-D ₃		
	4,6-androstadien-3,17- dione (6-D)	633-34-1	10.51	Testosterone-D ₃		
	dihydrotestosterone (DHT)	521-18-6	10.51	Testosterone-D ₃		
	4-androsten-3,17-dione	63-05-8	10.69	Testosterone-D ₃		
	Δ6-testosterone	2484-30-2	10.75	Testosterone-D ₃		
	testosterone (T)	58-22-0	10.92	Testosterone-D ₃		
	7α-hydroxytestosterone	62-83-9	11.24	Testosterone-D ₃		
	7β-hydroxy- dehydroepiandrosterone (7β-OH-DHEA)	2487-48-1	11.98	Testosterone-D ₃		
	formestane	566-48-3	13.14	Testosterone-D ₃		
4-hydroxytestosterone	2141-17-5	13.31	Testosterone-D ₃			
16α-hydroxyandrostene- 3,17-dione	63-02-5	13.60	Testosterone-D ₃			
Mix II	androsterone (Andro)	53-41-8	9.05	17α-methyl-testosterone		
	etiocholanolone (Etio)	53-42-9	9.18	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. Working mixtures, target analytes, CAS ID numbers, GC-MS retention times, internal standard used for quantification and concentration levels of the target analytes used to build the calibration curves.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Target Analyte	Calibration Range (ng/mL)	Weight	Model	Equation	F-test for heteroscedasticity	Levene Test for heteroscedasticity	Partial F-test for quadraticity	ANOVA - LoF	Squared Correlation Coefficient
T	2-125	1/x ²	Quadratic	-0.038x ² + 1.25x + 0.005	1.3 × 10⁻¹¹	1.4 × 10⁻⁵	6.8 × 10⁻⁸	5.1 × 10⁻³	0.9942
6-D	2-125	1/x	Quadratic	-0.020x ² + 0.599x + 0.023	2.6 × 10⁻⁶	1.7 × 10⁻⁵	1.7 × 10⁻³	1.8 × 10 ⁻¹	0.9949
Andro	100-2250	1/x ²	Linear	0.549x + 0.018	2.3 × 10⁻⁹	4.1 × 10⁻⁵	9.7 × 10 ⁻¹	1.4 × 10 ⁻¹	0.9999
16α-hydroxyandrost-3,17-dione	2-125	1/x ²	Quadratic	-0.013x ² + 0.164x + 0.006	2.0 × 10⁻¹¹	7.2 × 10⁻⁸	2.9 × 10⁻²	6.9 × 10 ⁻¹	0.9884
4-androsten-3,17-dione	2-125	1/x	Quadratic	-0.001x ² + 0.017x + 0.010	1.2 × 10⁻²	5.6 × 10 ⁻¹	3.5 × 10⁻³	8.0 × 10⁻⁴	0.9538
4-hydroxytestosterone	2-125	1/x ²	Linear	0.024x + 0.002	1.2 × 10⁻¹⁰	1.1 × 10 ⁻¹	1.5 × 10 ⁻¹	7.2 × 10 ⁻¹	0.9984
5α-adiol	2-125	1/x ²	Quadratic	-0.022x ² + 0.378x + 0.005	1.9 × 10⁻¹¹	5.8 × 10⁻⁴	6.6 × 10⁻⁵	4.8 × 10⁻²	0.9893
5β-adiol	2-125	1/x ²	Quadratic	-0.020x ² + 0.319x + 0.005	1.5 × 10⁻¹¹	2.6 × 10⁻⁴	1.1 × 10⁻⁵	2.4 × 10⁻²	0.9865
5β-androstan-3,17-dione	2-125	1/x ²	Quadratic	-0.045x ² + 0.676x + 0.059	4.7 × 10⁻¹⁰	8.4 × 10⁻⁹	2.5 × 10⁻²	1.8 × 10⁻⁴	0.9624
5-androstendiol	2-125	1/x	Quadratic	-0.024x ² + 0.347x + 0.022	1.6 × 10⁻⁴	3.5 × 10⁻⁵	2.1 × 10⁻⁵	2.1 × 10⁻⁶	0.9607
7α-hydroxytestosterone	2-125	1/x ²	Linear	0.171x + 0.011	5.8 × 10⁻⁸	6.8 × 10⁻³	8.2 × 10 ⁻¹	1.2 × 10 ⁻¹	0.9935
7β-OH-DHEA	2-125	1/x ²	Linear	0.286x + 0.035	2.2 × 10⁻⁹	3.7 × 10⁻⁴	5.6 × 10 ⁻¹	6.1 × 10 ⁻¹	0.9964
Δ6-testosterone	2-125	1/x ²	Quadratic	-0.013x ² + 0.577x + 0.032	3.5 × 10⁻⁵	7.1 × 10⁻³	6.2 × 10⁻³	7.8 × 10 ⁻²	0.9957
DHEA	2-125	1/x ²	Quadratic	-0.020x ² + 0.453x + 0.001	1.4 × 10⁻⁹	2.3 × 10⁻³	2.1 × 10⁻⁷	3.7 × 10⁻³	0.9933
DHT	2-125	1/x ²	Quadratic	-0.019x ² + 0.286x -	3.2 × 10⁻¹¹	6.9 × 10⁻⁶	2.8 × 10⁻⁴	1.2 × 10 ⁻¹	0.9922

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

				0.001						
E	2-125	1/x ²	Quadratic	-0.042x ² + 1.260x + 0.002	2.9 × 10⁻¹³	4.6 × 10⁻⁷	3.5 × 10⁻¹²	6.0 × 10⁻³	0.9946	
Etio	100-2250	1/x ²	Linear	0.472x + 0.031	6.2 × 10⁻⁹	1.0 × 10⁻³	5.7 × 10 ⁻¹	3.1 × 10 ⁻¹	0.9996	
formestane	2-125	1/x ²	Linear	0.748x + 0.074	6.3 × 10⁻⁴	3.0 × 10⁻⁴	9.2 × 10 ⁻²	8.4 × 10⁻³	0.9929	

Table 2. Calibration model parameters for all targeted analytes. Values of $p < 5 \times 10^{-2}$ (0.05) indicated the occurrence of heteroscedasticity (F-test and Levene test for heteroscedasticity), a relevant contribution of the quadratic term (partial F-test for quadraticity), and a significant deviation from the calibration model (ANOVA-LoF),. The corresponding values are reported in bold type.

Target Analyte	Calibration levels (Deviation %)					
	1	2	3	4	5	6
T	0.4	0.3	3.2	1.2	5.8	2.3
6-D	18.8	11.8	10.4	0.4	4.6	0.8
Andro	0.5	1.0	0.5	1.6	0.1	0.7
16α-hydroxyandrost-3,17-dione	3.2	11.9	7.7	5.1	10.7	13.0
4-androst-3,17-dione	9.3	15.5	12.1	11.0	6.9	2.9
4-hydroxytestosterone	7.4	16.2	8.2	10.3	9.8	16.4
5α-adiol	0.1	0.4	0.2	0.9	0.6	0.1
5β-adiol	0.4	1.7	1.3	0.2	0.2	0.1
5β-androstan-3,17-dione	21.7	8.4	12.2	13.5	15.9	3.6
5-androstendiol	37.3	22.2	20.3	5.4	14.3	3.9
7α-hydroxytestosterone	5.1	10.0	8.6	3.2	14.4	3.9
7β-OH-DHEA	16.4	8.4	13.6	4.5	1.8	0.4
Δ6-testosterone	17.8	6.3	13.8	3.5	6.9	1.0
DHEA	0.6	1.1	1.5	1.2	1.9	1.2
DHT	0.1	1.9	0.3	9.9	5.0	12.7
E	0.8	1.7	1.4	0.1	4.0	1.9
Etio	0.5	0.5	1.1	2.4	2.0	1.6
formestane	23.1	21.4	13.3	1.8	14.2	4.5

Table 3. Back-calculation results

Target Analyte	LOD (ng/mL) 6 points	LOD (ng/mL) 5 points	LOD (ng/mL) 4 points	Experimentally verified LOD (ng/mL)
T	0.7*	0.2	0.1	0.5
6-D	1.7*	0.5	2.3*	0.5
Andro	0.6	0.4	0.3	0.5
16α-hydroxyandrostene-3,17-dione	0.7*	0.8	0.7	1.0
4-androstene-3,17-dione	3.1*	3.3	0.8	1.0
4-hydroxytestosterone	0.7	0.6*	0.6*	1.0
5α-adiol	0.7*	0.5	0.3	0.5
5β-adiol	0.9*	0.3	0.2	0.5
5β-androstan-3,17-dione	4.5*	1.1*	0.5	0.5
5-androstene-3,17-diol	5.3*	1.1	1.8*	1.0
7α-hydroxytestosterone	1.8	1.8*	0.5	0.5
7β-OH-DHEA	0.5*	0.4	2.2*	0.5
Δ6-testosterone	1.0*	0.4	0.3	0.5
DHEA	1.2*	0.2	<0.1	0.5
DHT	0.5	0.4	<0.1	0.5
E	0.5*	0.1	<0.1	0.5
Etio	0.7	0.5	0.4	0.5
formestane	0.9	1.8*	0.4	0.5

*Although a quadratic trend was detected, it was ignored to compute the LOD values.

Table 4. The first three columns report the LOD values computed with the Hubaux-Vos algorithm using 6, 5, and 4 calibration levels. Weighting corrections were applied as described in Section 2.3.1. The last column displays the LOD values experimentally verified by spiking blank urine with the concentrations reported.

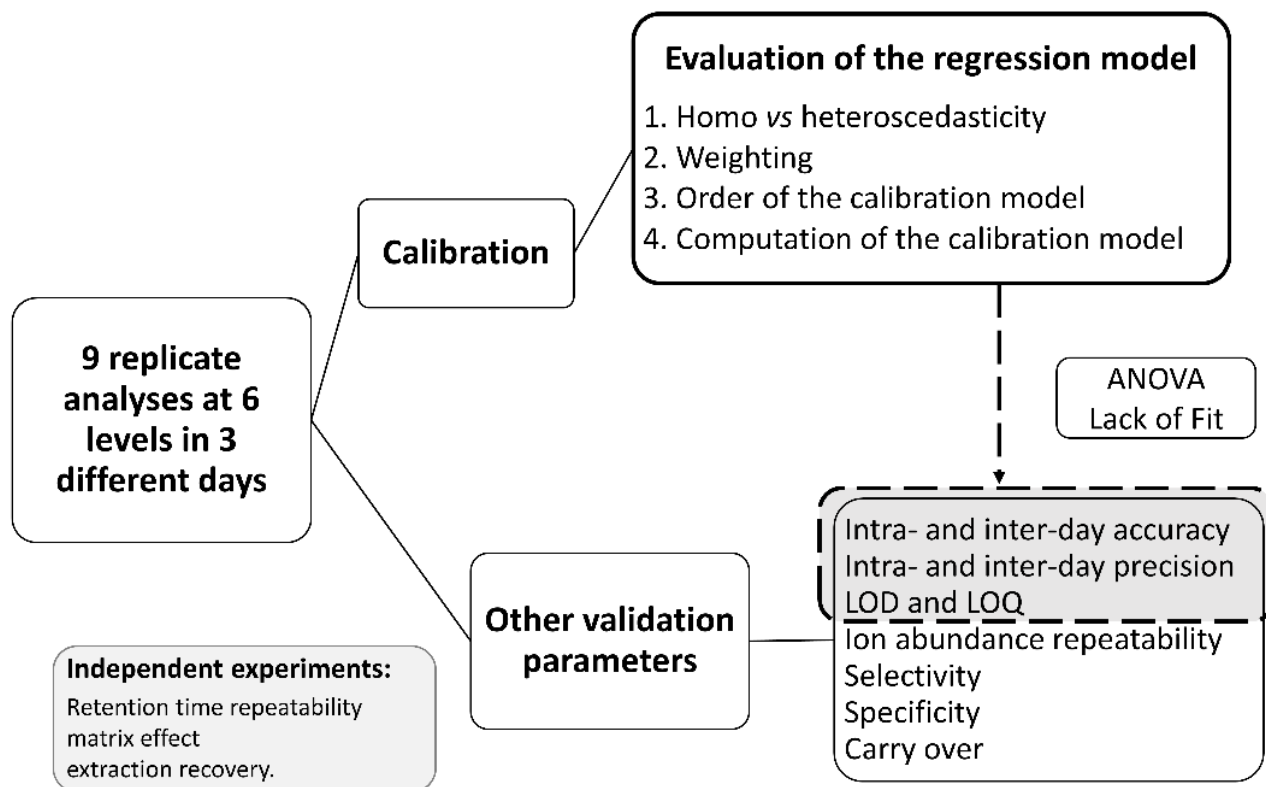
Target Analyte		Calibration Level (% bias)					
		1	2	3	4	5	6
T	Intra-day	0.4	0.5	-3.3	-1.2	5.6	-2.1
	Inter-day	0.4	0.3	-3.3	-1.3	5.4	-2.7
6-D	Intra-day	<i>16.1</i>	-12.2	-10.3	0.9	4.8	-10.3
	Inter-day	8.1	-12.3	-8.1	3.2	7.6	-1.5
Andro	Intra-day	0.2	-1.9	0.9	-0.8	-4.9	-5.1
	Inter-day	1.2	-0.3	0.6	2.3	0.8	-0.1
16 α -hydroxyandrostene-3,17-dione	Intra-day	-25.0	14.5	-6.0	10.9	-11.5	<i>-19.5</i>
	Inter-day	-8.9	10.7	-6.5	7.4	-3.7	-0.6
4-androsten-3,17-dione	Intra-day	-4.8	<i>-19.9</i>	21.65	-1.1	<i>18.0</i>	-23.1
	Inter-day	<i>-17.3</i>	-24.9	-14.1	-21.6	-21.9	-21.5
4-hydroxytestosterone	Intra-day	0.3	14.6	8.4	-1.1	<i>17.5</i>	-1.1
	Inter-day	-8.3	<i>16.3</i>	9.9	-9.4	11.4	<i>-19.9</i>
5 α -adiol	Intra-day	-4.9	-3.4	-3.7	-1.4	-4.7	-8.9
	Inter-day	1.3	0.6	1.2	2.0	1.0	-21.8
5 β -adiol	Intra-day	0.5	-1.4	1.3	0.9	-0.1	0.4
	Inter-day	1.2	-0.6	2.9	1.6	2.1	-20.6
5 β -androstan-3,17-dione	Intra-day	12.7	<i>-16.7</i>	-21.5	-4.0	20.3	-23.1
	Inter-day	14.9	-6.6	-9.1	-9.3	14.6	-12.2
5-androsten-3,17-diol	Intra-day	12.6	-23.6	<i>-16.7</i>	1.6	<i>16.5</i>	-6.4
	Inter-day	-11.5	-14.4	-23.6	<i>-15.1</i>	7.5	-20.7
7 α -hydroxytestosterone	Intra-day	4.8	-13.6	-14.6	-0.1	10.2	-9.4
	Inter-day	<i>-19.3</i>	<i>-18.3</i>	<i>-18.0</i>	2.7	-21.5	-25.0
7 β -OH-DHEA	Intra-day	8.4	-2.7	-9.2	<i>17.4</i>	10.6	-0.4
	Inter-day	9.0	-9.0	-12.5	8.3	4.0	5.0
Δ 6-testosterone	Intra-day	9.0	-11.9	-9.8	-1.7	5.9	-5.6
	Inter-day	11.2	-4.8	-12.5	-0.1	8.8	9.5
DHEA	Intra-day	1.4	-0.9	-1.7	1.0	1.6	-0.1
	Inter-day	0.8	-0.9	-1.2	1.4	1.9	-1.2
DHT	Intra-day	7.4	-11.0	-11.8	-11.6	-8.2	<i>-18.5</i>
	Inter-day	11.0	5.5	5.5	12.4	7.7	-9.2
E	Intra-day	5.4	0.1	-0.6	-0.1	3.2	-3.6
	Inter-day	0.8	-1.7	-1.4	0.0	3.9	-1.5
Etio	Intra-day	-0.2	0.3	1.3	-1.9	-5.1	-4.1
	Inter-day	0.8	1.8	2.6	3.5	-0.7	-0.6
formestane	Intra-day	25.0	-21.9	<i>-15.7</i>	-2.1	9.5	-3.2
	Inter-day	<i>18.8</i>	<i>-17.9</i>	<i>-18.6</i>	-1.6	8.8	-9.0

Table 5. Intra-day and inter-day accuracy results expressed in terms of bias. Good results are expected in the range $\pm 15\%$, and acceptable results in the range $\pm 20\%$ (reported in italics). Results exceeding 20 are indicated in bold.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Target Analyte		Calibration Level (CV %)					
		1	2	3	4	5	6
T	Intra-day	4	8	2	2	5	3
	Inter-day	4	8	2	4	5	4
6-D	Intra-day	39	21	12	8	10	4
	Inter-day	39	24	15	9	17	8
Andro	Intra-day	3	4	6	7	5	2
	Inter-day	10	11	18	15	12	8
E	Intra-day	4	3	5	2	3	2
	Inter-day	4	3	6	2	3	5
Etio	Intra-day	3	4	6	7	2	2
	Inter-day	13	14	20	16	15	10
5α-adiol	Intra-day	0.1	0.1	1	1	0.1	3
	Inter-day	0.2	0.4	2	2	3	26
5β-adiol	Intra-day	30	16	8	12	15	5
	Inter-day	16	13	19	16	16	23

Table 6. Intra-day and inter-day precision, expressed in terms of CV. Acceptable results are expected in the range $\pm 30\%$



29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure 1. Scheme of the validation protocol.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

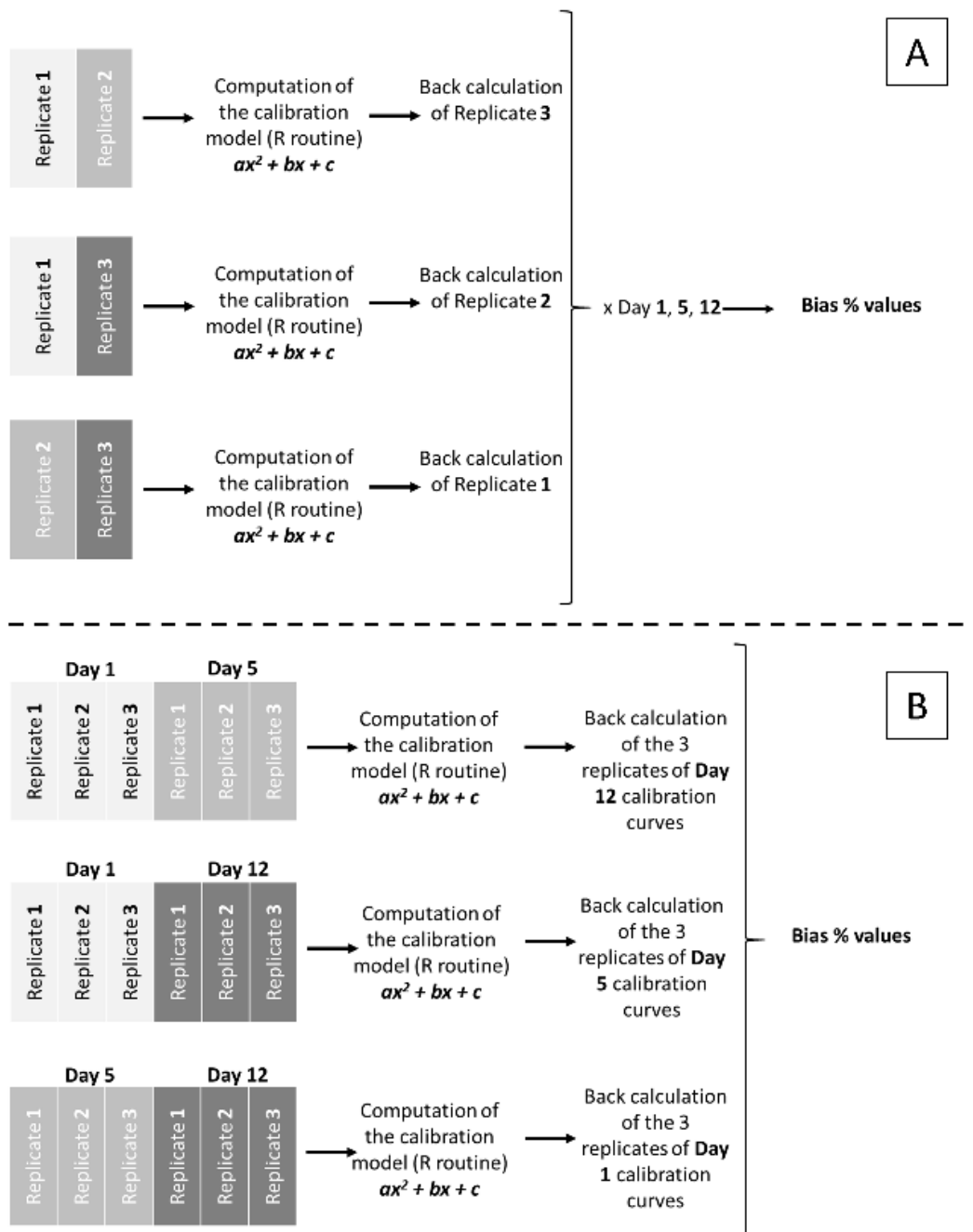


Figure 2. Operating scheme for the computation of the intra-day (A) and inter-day (B) accuracy.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

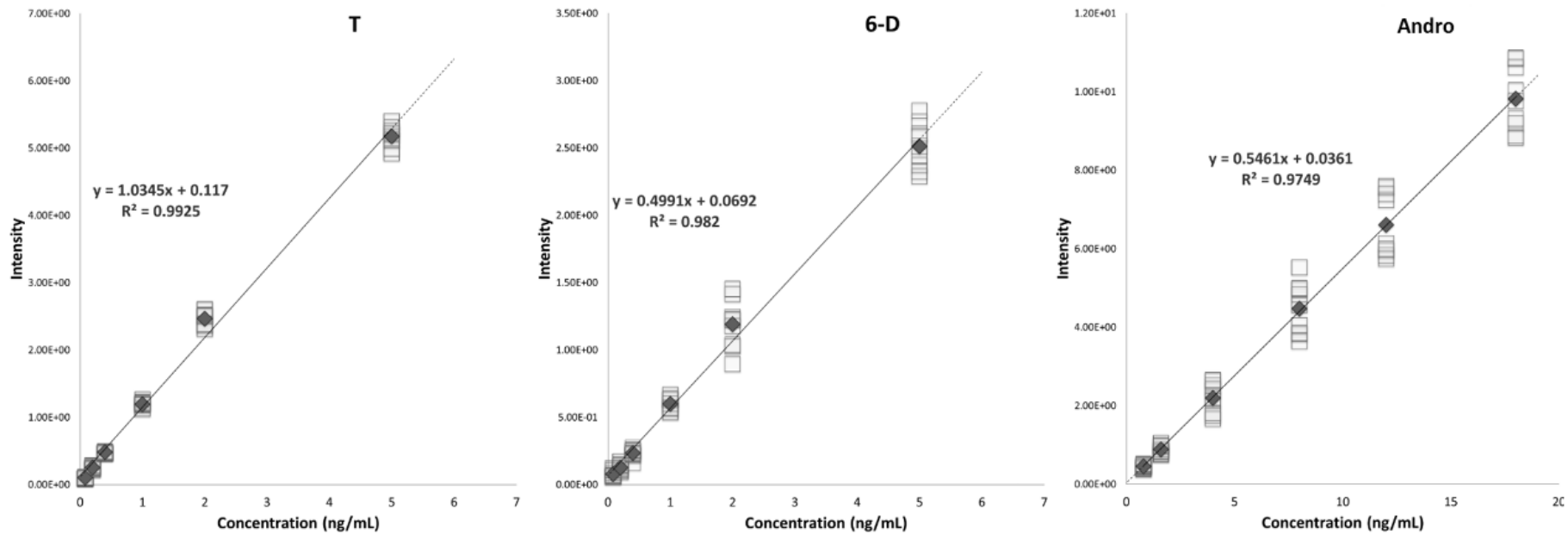


Figure 3. Distribution of the 54 calibration data-points (9 replicates × 6 calibration levels) for testosterone (T), 4,6-androstadien-3,17-dione (6-D), and androsterone (Andro).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

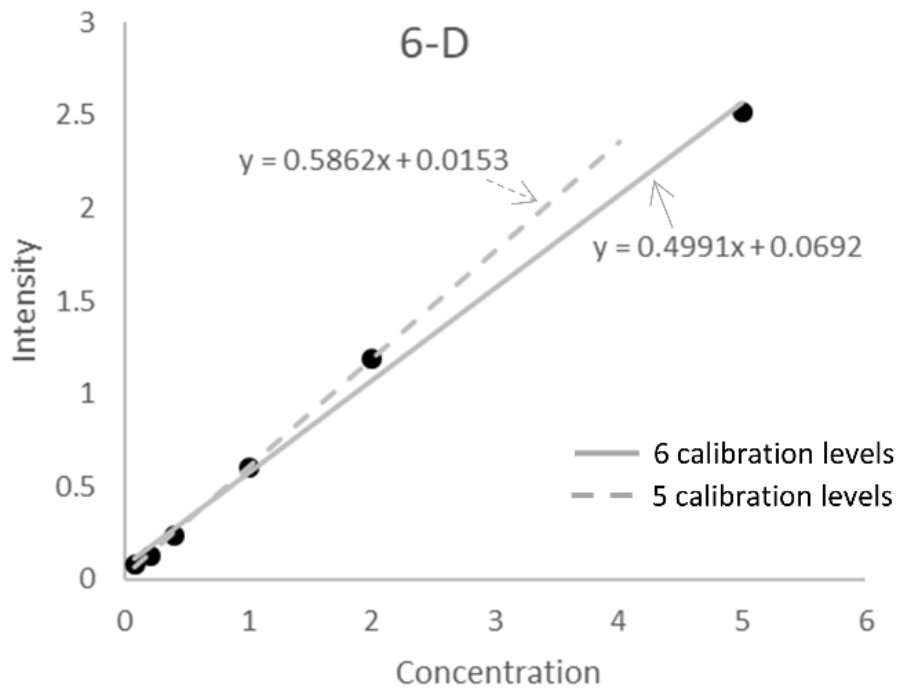
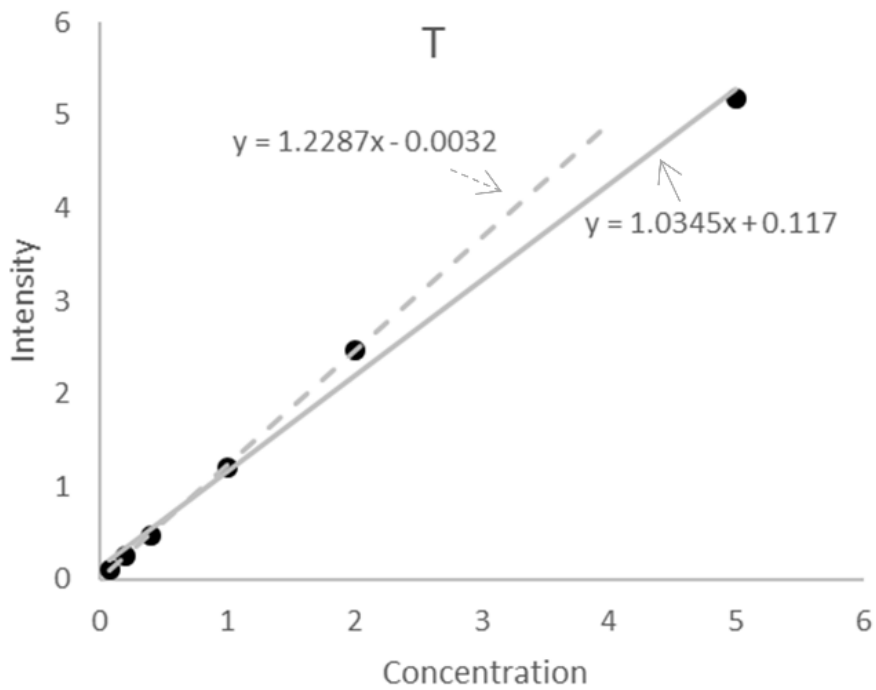


Figure 4. Linear regression for the six-points (solid line) and five-points (dashed line) calibration curves.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Supplementary Material

[Click here to download Electronic Supplementary Material \(online publication only\): Supplementary Material.docx](#)

Method Details (MethodsX)

[Click here to download Method Details \(MethodsX\): MethodsX.zip](#)