

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Application of multivariate statistics to the Steroidal Module of the Athlete Biological Passport: A proof of concept study

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1562338> since 2016-09-12T11:04:53Z

Published version:

DOI:10.1016/j.aca.2016.03.051

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)

Accepted Manuscript

Application of multivariate statistics to the Steroidal Module of the Athlete Biological Passport: a proof of concept study

Eugenio Alladio, Roberto Caruso, MS, Enrico Gerace, Ph.D., Eleonora Amante, MS, Alberto Salomone, Ph.D., Marco Vincenti, Prof.



PII: S0003-2670(16)30401-9

DOI: [10.1016/j.aca.2016.03.051](https://doi.org/10.1016/j.aca.2016.03.051)

Reference: ACA 234501

To appear in: *Analytica Chimica Acta*

Received Date: 26 November 2015

Revised Date: 24 February 2016

Accepted Date: 29 March 2016

Please cite this article as: E. Alladio, R. Caruso, E. Gerace, E. Amante, A. Salomone, M. Vincenti, Application of multivariate statistics to the Steroidal Module of the Athlete Biological Passport: a proof of concept study, *Analytica Chimica Acta* (2016), doi: 10.1016/j.aca.2016.03.051.

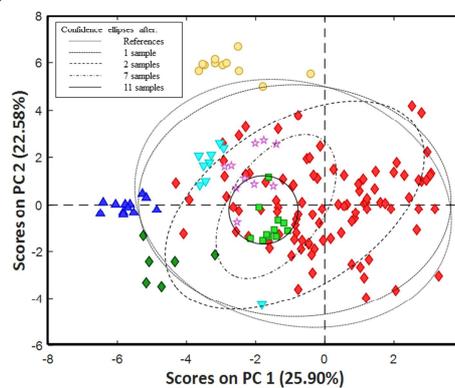
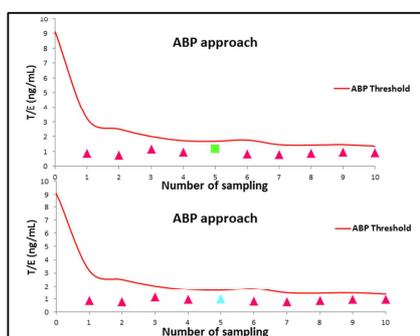
This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Application of multivariate statistics to the Steroidal Module of the Athlete Biological Passport

From univariate...



...to multivariate



ACCEPTED MANUSCRIPT

**Application of multivariate statistics to the Steroidal Module of the Athlete
Biological Passport: a proof of concept study.**

Eugenio Alladio^{1,2*}, Roberto Caruso², Enrico Gerace¹, Eleonora Amante², Alberto Salomone¹, Marco Vincenti^{1,2}

¹ *Centro Regionale Antidoping e di Tossicologia "A. Bertinaria", Regione Gonzole 10/1, 10043 Orbassano, Torino, Italy.*

² *Dipartimento di Chimica, Università degli Studi di Torino, Via P. Giuria 7, 10125 Torino, Italy*

***corresponding author:**

Eugenio Alladio, M.Sc.

Centro Regionale Antidoping e di Tossicologia "A. Bertinaria"

Regione Gonzole 10/1,

10043, Orbassano, Torino, Italy

Dipartimento di Chimica, Università degli Studi di Torino

Via Pietro Giuria 7,

10125 Torino, Italy

Tel.: +393460171979

E-mail: ealladio@unito.it; eugenio.alladio@gmail.com

email:

Roberto Caruso, MS: roberto.caruso@edu.unito.it;

Enrico Gerace, Ph.D.: enrico.gerace@antidoping.piemonte.it;

Eleonora Amante, MS: eleonora.amante@edu.unito.it;

Alberto Salomone, Ph.D.: alberto.salomone@antidoping.piemonte.it;

Marco Vincenti, Prof: marco.vincenti@unito.it.

Abstract

The Technical Document TD2014EAAS was drafted by the World Anti-Doping Agency (WADA) in order to fight the spread of endogenous anabolic androgenic steroids (EAAS) misuse in several sport disciplines. In particular, adoption of the so-called Athlete Biological Passport (ABP) – Steroidal Module allowed control laboratories to identify anomalous EAAS concentrations within the athletes' physiological urinary steroidal profile. Gas chromatography (GC) combined with mass spectrometry (MS), indicated by WADA as an appropriate technique to detect urinary EAAS, was utilized in the present study to develop and fully-validate an analytical method for the determination of all EAAS markers specified in TD2014EAAS, plus two further markers hypothetically useful to reveal microbial degradation of the sample. In particular, testosterone, epitestosterone, androsterone, etiocholanolone, 5α -androstane- $3\alpha,17\beta$ -diol, 5β -androstane- $3\alpha,17\beta$ -diol, dehydroepiandrosterone, 5α -dihydrotestosterone, were included in the analytical method. Afterwards, the multi-parametric feature of ABP profile was exploited to develop a robust approach for the detection of EAAS misuse, based on multivariate statistical analysis. In particular, Principal Component Analysis (PCA) was combined with Hotelling T^2 tests to explore the EAAS data obtained from 60 sequential urine samples collected from six volunteers, in comparison with a reference population of single urine samples collected from 96 volunteers. The new approach proved capable of identifying anomalous results, including (i) the recognition of samples extraneous to each of the individual urine series and (ii) the discrimination of the urine samples collected from individuals to whom "endogenous" steroids had been administered with respect to the rest of the samples population. The proof-of-concept results presented in this study will need further extension and validation on a population of sport professionals.

Keywords

Multivariate statistics; Athlete Biological Passport; Endogenous anabolic androgenic steroids; Principal Components Analysis; Hotelling's T^2 test; Gas chromatography–mass spectrometry.

1. Introduction

The recent developments in the fields of chromatography and mass spectrometry allowed toxicological laboratories to effectively detect synthetic anabolic steroids in urine samples [1]. As a matter of fact, screening methods are designed to monitor also several natural (endogenous) steroids, so as to identify Anti-Doping Rule Violations (ADRV) whenever intake of either class of steroids occurred. Traditionally, World Anti-Doping Agency (WADA) indicated specific endogenous steroids to be determined and compared to a reference population in terms of concentration ranges and abundance ratios [2]. Doped athletes taking extra-physiological doses of endogenous anabolic androgenic steroids (EAAS) were originally identified by their abnormal values of testosterone/epitestosterone (T/E) ratio [3–7], as compared to common physiological values and cut-off T/E thresholds established by WADA [2,6]. The detection of an abnormal T/E ratio value is reported as an atypical finding and confirmatory analysis by isotopic-ratio mass spectrometry (IRMS) is requested to unequivocally recognize the exogenous administration of doping agents [8].

The exogenous administration of EAAS for doping purposes remains problematic to detect, since a clear distinction between endogenous and exogenous origin of EAAS found in urine samples is still challenging for anti-doping laboratories [4,9–11]. As remarked by a recent WADA report [12], anabolic agents (including EAAS) represent the most abundant misused class of substances in elite sports. Moreover, anabolic effects are currently obtained also with low doses of EAAS mixtures [13]. In these cases, the wide inter-individual variability of EAAS levels does not lead to straightforward identification of drug misuse and confuses the interpretation of individual steroid profiles. For this purpose, WADA composed the Technical Document TD2014EAAS [8] with the aim of identifying anomalous EAAS urine concentrations within the athletes' individual physiological steroidal profile [14]; in particular, this document describes WADA's adoption of the Athlete Biological Passport (ABP) – Steroidal Module, that consists in monitoring the athletes

regularly over time by checking the possible occurrence of significant changes of direct markers concentrations within their steroidal urinary profile.

The ABP approach exploits an adaptive model based on Bayesian inference [3,6–8,14] that recognizes anomalous EAAS values possibly induced by ADRVs, by combining the variability measured from a reference population with the series of EAAS values progressively collected from the individual under examination (prior odds). After a few samplings, ABP compliance intervals progressively shift from a distribution based on reference populations to a personalized distribution, based on the previous concentration levels measured for the same individual. Atypical passport findings (ATP) are notified whenever EAAS values exceed the individualized thresholds, whose reduced width enhances the ABP sensibility toward the recognition of ADRV [9]. However, ABP does not provide a comprehensive evaluation of the detected EAAS concentration levels, since the adaptive Bayesian approach is applied to the longitudinal monitoring of single markers (i.e., testosterone) or single ratios of two markers (i.e., T/E), without any combined multivariate strategy. Under such circumstances, limited variations of EAAS values, resulting for example from urine manipulations or replacements (e.g., where the urine of a subject different from the tested one is misleadingly provided), could hardly be evidenced from single steroid profiling, since the substitute sample is expected to originate from a “clean” healthy individual, likely yielding physiological EAAS values within the expected compliance intervals.

Various techniques of multivariate statistics may assist control laboratories in combining the information provided by each EAAS marker together, resulting in comprehensive conclusions about the steroidal profile compliance of athletes. A few studies and feasible approaches have been already reported in the literature about the evaluation of multivariate methodologies [15,16] and the development of classification/discrimination models within the field of anabolic androgenic steroids and doping controls [17–19].

In the present study, gas chromatography combined with mass spectrometry (GC-MS) was used to detect the six EAAS involved in the ABP – steroidal module, including testosterone (T), epitestosterone (E), androsterone (A), etiocholanolone (Etio), 5 α -androstane-3 α ,17 β -diol (5a-Adiol), 5 β -androstane-3 α ,17 β -diol (5b-Adiol), plus two further EAAS, namely dehydroepiandrosterone (DHEA) and 5 α -dihydrotestosterone (DHT), and two markers of microbial degradation (5 β -androstenedione and 4-androstenedione). Moreover, urinary concentration ratios (e.g. T/E, A/T, A/Etio, 5 α -diol/5 β -diol and 5 α -diol/E) were monitored [8]. The quantitative GC-MS method was fully validated, and applied to the urine samples collected from a reference population of healthy subjects, who did not take any pharmaceutical drug potentially able to influence their steroidal profile. Subsequently, we developed a multivariate data analysis strategy that could discriminate from one another the urine samples of six individuals longitudinally-monitored during 40 days, even if they were consistent with the same reference population. Lastly, the same statistical method based on principal component analysis (PCA) [20,21] and Hotelling's T^2 tests was applied to the urine samples collected from individuals suffering from hormonal imbalance and taking Androderm® and Testovis® pharmaceutical drugs containing testosterone, who may represent a preliminary prototype for evaluating our procedure.

2. Materials and methods

2.1. Chemicals and reagents

Testosterone, epitestosterone, androsterone, etiocholanolone, 5 α -androstane-3 α ,17 β -diol, 5 β -androstane-3 α ,17 β -diol, dehydroepiandrosterone, 5 β -androstenedione, 4-androstenedione and testosterone-d3 were purchased from Steraloids Inc. (Newport, RI, USA); 5 α -dihydrotestosterone was provided by LGC Promochem SRL (Milan, Italy). Isopropyl alcohol, methanol, methyl tert-butyl ether (TBME), ethyl acetate, 17 α -methyltestosterone, dithioerythritol and N-Methyl-N-

(trimethylsilyl) trifluoroacetamide (MSTFA) were provided by Sigma-Aldrich (Milan, Italy). β -glucuronidase from *Escherichia coli* was purchased from Roche Life Science (Indianapolis, IN, USA) and ammonium iodide, was from TCI Co., Ltd. (Tokyo, Japan). C-18 endcapped Solid-Phase Extraction (SPE) cartridges were provided by UCT Technologies (Bristol, PA, USA). Ultra-pure water was obtained from a Milli-Q® UF- Plus apparatus (Millipore, Bedford, MA, USA). Stock standard solutions were stored at $-20\text{ }^{\circ}\text{C}$ until used.. Two working solution mixtures were prepared in methanol at final concentrations of respectively $1\text{ }\mu\text{g mL}^{-1}$ (working solution A, composed by all reference substances with the exception of androsterone and etiocholanolone) and $10\text{ }\mu\text{g mL}^{-1}$ (working solution B, composed by androsterone and etiocholanolone). These concentrations were chosen according to the upper limit of their endogenous concentration interval [9,22]. Two separate internal standard working solutions were prepared in methanol for testosterone-d3 and 17α -methyltestosterone at the final concentration of $10\text{ }\mu\text{g mL}^{-1}$.

2.2. Sample preparation

The sample preparation was experimentally-designed and optimized on the basis of studies reported in the literature [1,23–25] for urinary anabolic steroids. The urine sample (6 mL) was fortified with both internal standard solutions including an isotopically-marked molecule (testosterone-d3) and 17α -methyltestosterone. The UCT C-18 endcapped solid-phase extraction (SPE) cartridge was washed with isopropyl alcohol, methanol and distilled water (6 mL) in sequence. Then, the SPE cartridge was loaded with the urine sample and washed with 6 mL of distilled water. Lastly, both EAAS and their glucuronides were eluted with 6 mL of methanol. The resulting solution was dried under nitrogen at $50\text{ }^{\circ}\text{C}$ and the residue was dissolved at pH 7 with 2 mL of a 0.1 M phosphate buffer. β -glucuronidase ($50\text{ }\mu\text{L}$) was subsequently added and the mixture was incubated at $55\text{ }^{\circ}\text{C}$ for 1 h. Once the hydrolysis was completed, the mixture was cooled to room temperature and 2 mL of

0.1 M carbonate buffer (pH 9) were added. Liquid–liquid extraction was performed with 5 mL of TBME, then, the sample was shaken in a multimixer for 10 min, and subjected to centrifugation at 2500 rpm for 5 min. The extraction process was repeated and the combined organic phases were transferred into a vial and dried under nitrogen at 70 °C. The dry residue was derivatized with 50 μ L MSTFA/NH₄I/dithioerythritol (1.000:2:4 v/w/w) solution for 40 min at 70 °C. A 1- μ L aliquot was injected into the GC/MS system in the splitless mode.

2.3. Instrumentation

GC separations were performed using an Agilent 6890N instrument (Agilent Technologies, Milan, Italy) equipped with a J&W Scientific HP-1, 17 m \times 0.2 mm (i.d.) \times 0.11 μ m (f.t.) capillary column. Helium was employed as the carrier gas at a constant pressure of 21.5 psi. The chromatographic run was experimentally-designed and optimized [26] on the basis of previous studies [1,10,27]. The GC oven temperature was initially set at 120 °C, then raised to 177 °C with a 70 °C/min heating rate and subsequently raised to 236 °C with a 5 °C/min gradient. Lastly, the oven temperature was raised to 315 °C with a 30 °C/min ramp and the final temperature was maintained for 3 min. The total run time was 18.25 min. The GC injector and transfer line were maintained at 280 °C. 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.

2.4. Method validation

The following validation parameters were investigated according to WADA requirements [28]: linearity range, selectivity, specificity, limit of detection (LOD), limit of quantitation (LOQ), trueness, intra- and inter-assay precision, repeatability, matrix effect, extraction recovery and carry-over.

Linearity was evaluated in the concentration range of 2.0–500.0 ng mL⁻¹ for testosterone, epitestosterone, 5 α -androstane-3 α ,17 β -diol, 5 β -androstane-3 α ,17 β -diol, dehydroepiandrosterone, 5 α -dihydrotestosterone, 5 β -androstanedione and 4-androstenedione; in details, two independent calibration ranges were tested as follows: 2.0-50.0 ng mL⁻¹ (2.0, 5.0, 10.0, 15.0, 25.0 and 50.0 ng mL⁻¹) and 10.0-500.0 ng mL⁻¹ (10.0, 25.0, 50.0, 125.0, 250.0 and 500.0 ng mL⁻¹). This approach was adopted in order to guarantee robust calibrations over an extended concentration range. Moreover, linearity was evaluated within the range of 100.0–5000.0 ng mL⁻¹ for androsterone and etiocholanolone in two calibration ranges, i.e. 100.0-1500.0 ng mL⁻¹ (100.0, 250.0, 500.0, 750.0, 1000.0 and 1500.0 ng mL⁻¹) and 500.0-5000.0 ng mL⁻¹ (500.0, 1000.0, 1500.0, 2250.0, 3500.0 and 5000.0 ng mL⁻¹). Testosterone-d3 and 17 α -methyltestosterone were used as internal standards. The linear calibration parameters were evaluated using the least squares regression method; the determination coefficient (R²) was observed and several significance tests were performed to evaluate linearity, including lack-of-fit tests, Analysis of Variance (ANOVA) test, Mandel's test (in order to evaluate whether the calibration is linear or a quadratic curve), homoscedasticity studies, evaluations of the relative standard deviation of the slope and the residual standard error, together with the analysis of the deviation from back-calculated concentrations. Moreover, residual plots and homoscedasticity parameter were examined and successfully tested.

Selectivity and specificity were determined by the analysis of ten blank deionized water samples spiked with all the target analytes at the second, fourth, and sixth concentration levels of both concentration ranges. The signal-to-noise ratio (S/N > 3) was measured on the selected-ion

chromatograms at the expected retention times of all the analytes of interest. Moreover, the presence of interfering compounds at the retention time of the target analytes was examined.

LOD and LOQ values were determined using the Hubaux-Vos' technique [29]; in particular, three independent calibration lines were prepared for all the target analytes and a significance level of 95% was selected at the corresponding number of degrees of freedom. Then, Hubaux-Vos' algorithms allowed us to calculate LOD and LOQ values. The calculated values were experimentally tested with analyte concentrations extremely close to the detectable and quantifiable values, respectively, confirming the correct estimation.

Trueness, intra- and inter-assay precision were evaluated on ten deionized water samples spiked at the concentration levels of 100.0, 1500.0 and 3500.0 ng mL⁻¹ for androsterone and ethiocolanolone, and 10.0, 125.0 and 250.0 ng mL⁻¹ concentration levels for the remaining target analytes. Trueness, intra- and inter-assay precision were estimated as CV% and percent bias, respectively; satisfactory results were expected to lie within ±15%.

Retention time repeatability was verified on 30 real urine samples together with blank water samples spiked at different concentration levels, namely the ones used for trueness, intra- and inter-assay precision evaluation. Retention times for the real urine samples belonging to the examined individuals were also monitored over extended periods of time. Deviations below 1% from calibrators and controls were considered acceptable. Ion abundance repeatability was evaluated on the selected qualifying-ion chromatogram for each target analyte. In particular, the variations of the selected ion intensity were considered satisfactory within ±20%, with respect to the control.

Matrix effect was assessed by comparing the experimental results from two sets of solutions [30] at two concentrations (100.0 and 3500.0 ng mL⁻¹ for androsterone and ethiocolanolone; 10.0 and 250.0 ng mL⁻¹ for the other target analytes). The first set was composed by blank urine samples (collected from a 5 months old female child); the second by blank deionized water samples; both

sets were spiked after the extraction step. Then, the matrix effect was calculated as the percentage ratio between the quantified concentration levels of the target analytes from the first set and the ones detected from the second set. The percentage difference showed matrix suppression (values below 100 %) or enhancement (values above 100 %).

Extraction recovery was assessed by comparing the experimental results from two sets of solutions [30] at two concentrations (100.0 and 3500.0 ng mL⁻¹ for androsterone and ethiocolanolone; 10.0 and 250.0 ng mL⁻¹ for the other target analytes). The first set was composed by blank urine samples spiked after the extraction step and the second set was composed by blank urine samples spiked before the extraction step. Extraction recovery was calculated by the ratio between the quantified concentration levels of the target analytes from the second set and the ones detected from the first set.

Carry-over effect was evaluated by injecting five distilled water extracts and five urine samples spiked with all the analytes at the highest concentrations in alternate sequence (5000.0 ng mL⁻¹ for androsterone and ethiocolanolone; 500.0 ng mL⁻¹ for the other target analytes). In particular, the signal to noise ratio had to be lower than 3 for each monitored ion chromatogram in order to consider carry-over effects negligible.

2.5. *Samples collection and data description*

Single urine samples were collected from 96 volunteers (80 men, 16 women, aged 18–40) to provide a reference population database. Further 60 urine samples were collected from 6 volunteers (5 men, 1 woman, aged 23–29) to evaluate the intra-individual variability of steroidal profiles during a 40 days period of time, with a 2 sampling/week frequency, and a 3-days minimum interval between two consecutive collections. The urine samples collected from volunteers, named S1–S6, were as follows: 6 samples from volunteer S1, 11 samples from S2, 13 samples from S3, 9 samples

from S4, 10 samples from S5 and 11 samples from S6. Moreover, urine samples were collected from 12 individuals suffering from hormonal imbalance and taking Androderm® and Testovis®, pharmaceutical drugs containing testosterone. All subjects provided a signed informed consent to donate urine. The collected samples were frozen at -20°C , and analyzed within the following 3 months.

2.6. Chemometrics

Multivariate data analysis was carried out on Matlab® (The MathWorks, MA, USA) version 7.13.0 with PLS_Toolbox version 8.0 [31]. An initial PCA model was built using the 96 single urine samples forming the reference population. A training set matrix was arranged, consisting of 96 rows (representing each subject) and 13 columns (representing 8 urinary steroids plus the 5 steroids ratios suggested by WADA, i.e. T/E, A/T, A/Etio, 5α -diol/ 5β -diol and 5α -diol/E [8]). The descriptive statistics relevant to the training set data matrix is reported in Table 1, including the minimum, inter-quartile ranges (IQ1–IQ3), median, mean and maximum concentration values. All data were autoscaled and a cross-validation procedure was performed by applying the venetian blinds design and a number of data splits equal to 5 [20]. The optimal number of principal components (PC) to build the PCA model was determined from the Predicted Residual Sums of Squares (PRESS) and Root Mean Squared Error of Cross-Validation (RMSECV) [20,21]. Parameters such as eigenvalues, percentage variance captured by each PC (V%) and percentage cumulative variance captured by the model (CV%) were evaluated too, together with Q residuals and Hotelling T^2 [20,21]. Once the PCA model was developed, concentration levels of urine samples from the 6 time-monitored volunteers (data matrixes for each volunteer are reported in Appendix A (Supplementary Material) were introduced into the model independently, in order to recognize any significant variation in the longitudinally-monitored steroidal profiles of the subjects under examination. For this purpose, multivariate Hotelling's T^2 test was adopted [32]; in particular, PCA scores values for each volunteer were extracted and then compared with the ones belonging to the corresponding time-

monitored subject by means of Hotelling's T^2 tests. This approach was performed to observe whether the new samples of a certain individual monitored over the time could be considered as normal (i.e., compatible with the calculated distribution), with respect to his/her previous physiological levels, rather than anomalous and, maybe, induced by an ADRV.. This test was applied on 6 volunteers' steroidal profiles and allowed to recognize any anomalous value. Hotelling's T^2 test was performed on R software version 3.2.1 [33] with *Hotelling* package [34].

3. Results and discussion

3.1. Method validation

Linear calibration was observed for all the target analytes within their specific calibration ranges; satisfactory squared correlation coefficients (R^2) were observed (Table A in the Supplementary Material) and all significance tests checking various calibration parameters were positively verified, as they showed no significant deviation from linearity. Moreover, homoscedasticity tests were successfully passed. Retention time precision, selectivity and specificity also proved to be satisfactory, and no interfering signals were detected at the retention times of the target analytes. LOD and LOQ values calculated with Hubaux-Vos' technique are reported in Table B of the Supplementary Material. Remarkably, the developed method provided LOQ values for testosterone and epitestosterone equal to 1.7 and 1.9 ng mL⁻¹, respectively, satisfying WADA requirements for these target analytes (LOQ ≤ 2.0 ng mL⁻¹). Trueness, intra- and inter-assay precision results turned out adequate too, as the percent bias and the CV% were lower than 15.0% at all tested concentration levels (Table C in the Supplementary Material). For each target analyte, matrix effect and extraction recovery results are shown in Table D in the Supplementary Material. Both parameters showed variations within ±20% for all the analytes under investigation. It was concluded that neither the

extraction recovery nor the matrix effect prevented the correct determination of the corresponding analytes, Lastly, absence of any carry-over effect was observed.

3.2. *Data summary*

A typical GC-MS chromatogram obtained from a real urine sample is reported in Figure 1. The mean results for both reference population (96 urine samples) and six volunteers S1-S6 (60 urine samples) are reported in Figure 2, while the complete data are reported in Table 1 and Appendix A, respectively. For some EAAS and ratios, WADA provides reference limits, which are denoted by red squares in Figure 2. For example, WADA's reference limit for T/E ratio is equal to 4.0, concentrations for T or E are equal to 200 ng mL^{-1} , concentrations for A and Etio are equal to $10.000 \text{ ng mL}^{-1}$, while A/Etio ratio has a lower limit of 0.4 for males and an upper limit of 4 for both males and females. All EAAS concentrations obtained from all subjects lied within the WADA's reference limits: although average values are reported in Figure 2 for simplicity, neither maximum nor minimum values for each EAAS and ratios exceeded WADA's reference limits. Moreover, concentration of T and E lower than 50 ng mL^{-1} were always found in females for both reference population and selected volunteers. CV% values were less than 30% for all the target analytes and ratios of longitudinally-monitored volunteers.

3.3. *Reference Population PCA model*

The 96×13 data matrix was adopted as the training set in order to develop a PCA model. From calculation, the RMSECV value equal to 0.79, together with CV% of 82.47%, indicated that the optimal number of PCs to be considered was five. Accordingly, also Q residuals and Hotelling T^2 criteria were satisfied. Scores plots reporting PC1 (V% = 25.90%) vs. PC2 (V% = 22.58%) and PC3

(V% = 16.08%) vs. PC4 (V% = 9.93%) are depicted in Figure 3A-3B. The single urine collection for each of 96 volunteers yielded no significant clusters nor subpopulations in the scores plots. Loading plots showing PC1 vs. PC2 and PC3 vs. PC4 are reported in Figure 3C-3D. No variables selection techniques were applied, making all the target analytes and their relative ratios equally contribute to the statistical analysis, as required by WADA for ABP.

3.4. Time-monitored volunteers: PCA models & Hotelling's T^2 tests

The concentration levels of 8 steroidal markers plus 5 urinary ratios were evaluated for 60 urine samples collected from 6 volunteers under longitudinal time-monitoring conditions for over one month. Six matrices (one for each volunteer; Appendix A) were prepared and used as evaluation sets; in particular, a 6×13 matrix was set for volunteer named S1, while the number of rows (samples) were 11 for both volunteer S2 and S6, 13 for S3, 9 for S4, and 10 for S6. These data matrices were translated into the new PCA space (using the loading matrix obtained from the reference population) and introduced into the PCA model independently from one another, in order to evaluate the scores distributions for each volunteer, and obtain an interpretation model as close as possible to ABP. PC1 vs. PC2 scores plots from the time-monitored volunteers are reported in Figure 4, in the presence (4A) or absence (4B) of the reference population. The longitudinal scores for each volunteer are designated by green diamonds (S1), green squares (S2), blue triangles (S3), light blue inverted triangles (S4), purple stars (S5), yellow circles (S6), and red diamonds (reference population). In considering Figure 4, it should be taken into account that only PC1 and PC2 are depicted, accounting for no more than 48.5% variance. Since also PC3, PC4, and PC5 proved to be significant in representing the data distribution, perception of data point distances in Figure 4 is somehow distorted.

Hotelling's T^2 tests were initially performed on the scores values of each volunteer in comparison with the ones of the reference population: no significant differences were observed among these groups. In fact, Hotelling's T^2 test results suggest that the scores of the collected samples are arranged within the multidimensional space occupied by the scores of the reference population individuals, indicating that their steroid profile is not remarkably different from the reference population. Moreover, within-group multivariate Hotelling's T^2 test was separately performed on the scores values of each volunteer: no significant difference was found within the steroidal profiles' variations of subjects S1, S2, S3, S5 and S6, as p-values ranged from 0.439 up to 0.8474. In contrast, Hotelling's T^2 test showed a significant p-value of 7.20×10^{-7} (i.e., much less than 0.05, corresponding to 95% significance level) when executed on male subject S4 scores. One sample turned out anomalous after the application of Hotelling's T^2 test, namely sample S4.5 marked by an arrow in Figure 4B. The original S4.5 urine sample was intentionally replaced with one belonging to a female individual during sampling operations. Despite the overall steroidal profile of S4.5 is appreciably different from the other S4 samples even without using statistical tools, it is still interesting that the PCA model, combined with multivariate Hotelling's T^2 test, provides a quantitative evaluation of the dissimilarity of this specific steroidal profile with respect to the others, since it uses the p-value based on the whole information provided by the extended panel of steroidal markers (Figure 4B). On the other hand, single steroidal markers may not show any significant variation from sample replacement, possibly making the application of the ABP approach unable to recognize the extraneous sample. For example, the T/E ratio for sample S4.5 did not exceed the Bayesian threshold calculated on the S4 longitudinal measurements (Figure 4C).

The same approach within the PCA model was applied to all collected samples from longitudinally-monitored volunteers, after removal of sample S4.5 from S4 series. Further Hotelling's T^2 analysis was performed on selected samples, in order to test the robustness of our method. First, all S5 PCA scores were compared with the closest scores relative to different volunteers, whose data points are

located nearby S5 data points in the PC1 vs PC2 Cartesian diagram reported in Figure 4D. Five urine samples were tested, namely samples S2.2, S2.4, and S2.10 from subjects S2, and S4.1 and S4.2 from subject S4 (Figure 4D). All S2, S4, and S5 are male individuals. Hotelling's T^2 tests were executed to compare S5 samples vs. S2 and S4 samples. In the multi-dimensional space of significant PC variables, our PCA model proved to distinguish the samples belonging to subject S5 from each one collected from S2 (p-values of 1.71×10^{-4} , 2.11×10^{-4} , and 7.82×10^{-5}) and S4 (p-values of 2.89×10^{-4} and 1.54×10^{-3}). Notably, S5 provides the most scattered data-points among the studied subjects, corresponding to the highest variance and the least discriminating conditions, but nevertheless all S2 and S4 samples were easily recognized as extraneous to the S5 series.

On the other hand, the S2 and S4 samples previously tested were introduced in the sequence of longitudinal S5 samplings, as envisaged in the traditional ABP protocol. The steroid markers' values for all S2 and S4 samples fell inside the expected limits of the monovariate adaptive Bayesian model built on the S5 series, as is depicted for the T/E ratio in samples S2.10 and S4.1 (Figures S1A-B in the Supplementary material). Again, it was not possible to recognize any sample extraneous to the S5 series using the monovariate ABP approach, since the values of their steroid markers invariably fell below the estimated Bayesian threshold build from the S5 longitudinal measurements. In contrast, the PCA model combined with Hotelling's T^2 tests proved to represent a useful tool to distinguish extraneous samples, anomalous marker values, or significant variations in the steroidal profiles of individuals.

Hotelling's T^2 tests were also utilized among all time-monitored volunteers (S1-S6), in order to evaluate the sensitivity and the specificity of the model. Again, the model proved satisfactory, as no false positives for S1-S6 (0%) and only two false negatives results for S3 (4%) were observed. In particular, false positive rate described the number of urine samples provided from a certain individuals that were classified as belonging to a different subject. Conversely, false negative rate reported the number of samples belonging to different individuals that were labelled as belonging to

the same subject. Calculated p-values ranged from a minimum of 1.11×10^{-16} to a maximum of 1.39×10^{-4} .

In order to expand the Bayesian adaptive model into a multivariate statistical pattern, multidimensional Bayesian boundaries were built within the PC space for each of the S1-S6 longitudinal samplings. To visualize these boundaries, confidence ellipses at a significance level of 95% were calculated and reported in Figures 5A-5B (limited to a two-dimensional space), which are relative to volunteers S2 and S3. These ellipses were drawn in PC1 vs PC2 scores plot in sequential refinement with respect to the progressive introduction of each single sample of the time-monitored volunteers. In particular, the first confidence ellipse was calculated on the basis of the original reference population. Then, a second ellipse was calculated after the introduction of the first sample from subject S2 (or S3). The same Bayesian approach, providing prior odds, was followed after the introduction of the second, the seventh, and the last sample for each investigated subject. As is evident in Figures 5A-5B, the narrowest ellipses perfectly enclose the distributions of the longitudinally-monitored samples, thus indicating a specific space limit in the scores plot where the samples related to the corresponding subjects are expected to be found. Any displacement out of these confidence ellipses for further samples subsequently analyzed highlights a suspected ATP sample, which requires further investigation in order to clear up an alleged ADRV case. S2 and S3 extended series (13 samples each) were selected as examples of distinct behaviors, the former showing a progressive focusing within narrower dimensions of the same space, whereas in S3 a shift toward low PC1 values is observed as long as new samples are added to the series. Notably, S3 is the only female volunteer, while the reference population has a 83% prevalence of male subjects, which explains the partial displacement of the S3 series. Nevertheless, in the five-dimensional space of PC1-PC5, no S3 samples would be classified as abnormal with respect to the initial reference population ellipse, unlike it misleadingly appears in the two-dimensional Figure 5B. Similar models can be obtained from shorter series, as it occurs for S1, with six data-points only.

3.5. Testosterone intake: PCA models & Hotelling's T^2 tests

The described multivariate approach was utilized to compare a new model built with the reference population together with S1-S6 urine samples with the samples collected from 12 individuals suffering from hormonal imbalance and taking either Androderm® or Testovis®. The steroid profile data of these individuals are shown in Appendix B (Supplementary Material). The corresponding PCA scores plot is reported in Figure 6A, where the urine samples relative to pharmacologically-treated subjects are denoted by green diamonds. Quite obviously, the PCA model combined with Hotelling's T^2 test effectively singled out the subjects taking testosterone, as their position within the multidimensional score space proved to be significantly different from any other population of untreated subjects (for example, the p-value relative to the comparison between the treated subjects and S5 samples was equal to 3.09×10^{-7}). In this case, however, also the monivariate ABP approach proved adequate to detect the anomalous steroid profile (Figure 6B) induced by Androderm® or Testovis® intake. For example, the mean T and T/E values for the treated subjects were 58.9 and 2.77, respectively, while the mean T and T/E values were 33.6 and 1.20 for the reference population. Thus, both multivariate data analysis and monivariate ABP equally represent efficient ways to detect anomalous steroid marker values associated to the use of pharmaceutical drugs containing testosterone.

4. Conclusions

At the current stage, the monivariate adaptive model utilized in the steroidal module of the ABP proved to be a useful tool for detecting EAAS misuse [35], even if many confounding factors may be considered in the evaluation of individual steroid profiling [36]. The presence of both

endogenous and exogenous influencing variables suggest extreme care in the application of purely statistical rules in the probabilistic assessment an ADRV [36].

The present concept study has been intended to verify and measure the potential advantages arising from the use of a simple multivariate statistical approach to the interpretation of data that are intrinsically multi-parametric and highly inter-correlated. More sophisticated statistical methods, including variable selection algorithms and partial least squares based methods, could be applied to the ABP interpretation in the future, once the seeming advantages of the multivariate approach had been verified on solid databases from real sport athletes.

The contribution offered by the present study conducted on a generic reference population without specific involvement in professional sport practice consists of several starting points to be sustained with experiment conducted under a variety of experimental conditions. First of all, the intake of EAAS-containing drugs for clinical purposes results in a macroscopically modified urinary steroid profile, which is clearly distinguishable from those of subjects not under pharmacological treatment. On a PCA scores plot, these samples are confined within a rather restricted area to form a cluster located at large geometrical distance from the reference population samples distribution. A forthcoming study will investigate how long, after the intake of the last EAAS dose, does this easy discrimination lasts.

A second notable evidence offered by the present study is that, on the multivariate scale provided by the steroid profile, a single urine sample collected from a certain individual can be almost unfaillingly be distinguished from a series of urine samples collected from a different subject. This conclusion represents a reasonable basis to check the suspect cases of urine replacement, and facilitate the decision process of recognizing this type of ADRV.

The third practical suggestion provided by this study is that the concepts of (i) Bayesian adaptive model of steroid profile and (ii) individualized threshold of steroid physiological variation find

straightforward extension into the multidimensional space of statistically significant Principal Components, possibly delimiting a tolerance hyperspace with concurrent higher confidence and selectivity. Likewise the current ABP, the tolerance threshold hyperspace progressively shrink as many urine samples are longitudinally collected from the same individual.

Acknowledgements

No specific financial support was received for this study. Continuous support from M.I.U.R. and Regione Piemonte is kindly acknowledged.

References

- [1] M.K. Parr, W. Schänzer, Detection of the misuse of steroids in doping control, *J. Steroid Biochem. Mol. Biol.* 121 (2010) 528–37.
- [2] WADA, Technical Document – TD2004EAAS, 2004.
- [3] P.-E. Sottas, N. Baume, C. Saudan, C. Schweizer, M. Kamber, M. Saugy, Bayesian detection of abnormal values in longitudinal biomarkers with an application to T/E ratio, *Biostat.* 8 (2007) 285–96.
- [4] WADA, Guidelines for the Reporting and Management of Elevated T/E Ratios and Endogenous, 2011.
- [5] N. Robinson, V. Castella, C. Saudan, P.-E. Sottas, C. Schweizer, N. Dimo-Simonin, et al., Elevated and similar urinary testosterone/epitestosterone ratio in all samples of a competition testing: suspicion of a manipulation, *Forensic Sci. Int.* 163 (2006) 148–51.
- [6] P.-E. Sottas, C. Saudan, C. Schweizer, N. Baume, P. Mangin, M. Saugy, From population- to subject-based limits of T/E ratio to detect testosterone abuse in elite sports, *Forensic Sci. Int.* 174 (2008) 166–72.
- [7] P. Van Renterghem, P. Van Eenoo, P.-E. Sottas, M. Saugy, F. Delbeke, A pilot study on subject-based comprehensive steroid profiling: novel biomarkers to detect testosterone misuse in sports, *Clin. Endocrinol.* (2011) 134–140.
- [8] World Anti-Doping Agency, WADA Technical Document – TD2014EAAS Endogenous Anabolic Androgenic Steroids, 2014.
- [9] P. Van Renterghem, P. Van Eenoo, F.T. Delbeke, Population based evaluation of a multi-parametric

- steroid profiling on administered endogenous steroids in single low dose, *Steroids*. 75 (2010) 1047–57.
- [10] P. Van Eenoo, W. Van Gansbeke, N. De Brabanter, K. Deventer, F.T. Delbeke, A fast, comprehensive screening method for doping agents in urine by gas chromatography-triple quadrupole mass spectrometry, *J. Chromatogr. A*. 1218 (2011) 3306–3316.
- [11] C. Saudan, N. Baume, N. Robinson, L. Avois, P. Mangin, M. Saugy, Testosterone and doping control, *Br. J. Sports Med.* 40 Suppl 1 (2006) i21–4.
- [12] World Anti-Doping Agency, 2012 Anti-Doping Testing Figures Report, 2012.
- [13] H. Geyer, W. Schanzer, M. Thevis, Anabolic agents: recent strategies for their detection and protection from inadvertent doping, *Br. J. Sports Med.* 48 (2014) 820–826.
- [14] P.-E. Sottas, M. Saugy, C. Saudan, Endogenous steroid profiling in the athlete biological passport, *Endocrinol. Metab. Clin. North Am.* 39 (2010) 59–73, viii–ix.
- [15] H.R. Norli, K. Esbensen, F. Westad, K.I. Birkeland, P. Hemmersbach, Chemometric evaluation of urinary steroid profiles in doping control, *J. Steroid Biochem. Mol. Biol.* 54 (1995) 83–88.
- [16] F. Ponzetto, F. Mehl, J. Boccard, N. Baume, S. Rudaz, M. Saugy, et al., Longitudinal monitoring of endogenous steroids in human serum by UHPLC-MS/MS as a tool to detect testosterone abuse in sports, *Anal Bioanal Chem.* 408 (2016) 705–719.
- [17] P. Van Renterghem, P.-E. Sottas, M. Saugy, P. Van Eenoo, Statistical discrimination of steroid profiles in doping control with support vector machines, *Anal. Chim. Acta.* 768 (2013) 41–8.
- [18] E.O. Cannon, A. Bender, D.S. Palmer, J.B.O. Mitchell, Chemoinformatics-based classification of prohibited substances employed for doping in sport, *J. Chem. Inf. Model.* 46 (2006) 2369–80.
- [19] P. Sottas, N. Robinson, S. Giraud, F. Taroni, Statistical Classification of Abnormal Blood Profiles in Athletes, *Int. J. Biostat.* 2 (2006) 1–21.
- [20] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part A*, first, Amsterdam, 1997.
- [21] R. Bro, A. Smilde, Principal component analysis, *Anal. Methods*. 6 (2014) 2812–2831.
- [22] 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) 154–63.
- [23] C. Shackleton, J. Whitney, Use of Sep-pak cartridges for urinary steroid extraction: evaluation of the method for use prior to gas chromatographic analysis, *Clin. Chim. Acta.* 107 (1980) 231–243.
- [24] P. Vestergaard, The hydrolysis of conjugated neutral steroids in urine, *Acta Endocrinol. Supp.* 217 (1978) 96–120.
- [25] M. Donike, K. Barwald, K. Klostermann, W. Schanzer, J. Zimmermann, The detection of exogenous testosterone, *Int. J. Sport Med.* 4 (1983) 68–168.
- [26] R. Leardi, Experimental design in chemistry: A tutorial, *Anal. Chim. Acta.* 652 (2009) 161–72.
- [27] P. Van Renterghem, P. Van Eenoo, W. Van Thuyne, H. Geyer, W. Schänzer, F.T. Delbeke, Validation of an extended method for the detection of the misuse of endogenous steroids in sports, including new hydroxylated metabolites, *J. Chromatogr. B.* 876 (2008) 225–35.
- [28] World Anti-Doping Agency, International standard for laboratories, version 6.0., 2010.
- [29] A. Hubaux, G. Vos, Decision and detection limits for calibration curves, *Anal. Chem.* 42 (1970) 849–855. <http://pubs.acs.org/doi/abs/10.1021/ac60290a013> (accessed March 14, 2014).
- [30] B. Matuszewski, Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS, *Anal. Chem.* 75 (2003) 3019–3030.

<http://pubs.acs.org/doi/abs/10.1021/ac020361s> (accessed March 14, 2014).

- [31] B.M. Wise, N.B. Gallagher, R. Bro, J.M. Shaver, W. Windig, R.S. Koch, PLS_Toolbox 8.0, Eigenvector Research, Inc. (2006).
- [32] G. Zadora, A. Martyna, D. Ramos, C. Aitken, Statistical Analysis in Forensic Science - Evidential Value of Multivariate Physicochemical Data, First, John Wiley & Sons, Ltd., Chicester, UK, 2014.
- [33] R Core Team, R: A language and environment for statistical computing, (2014). <http://www.r-project.org/>.
- [34] J.M. Curran, Hotelling: Hotelling's T-squared test and variants, (2013). <http://cran.r-project.org/package=Hotelling>.
- [35] N. Baume, H. Geyer, M. Vouillamoz, R. Grisdale, M. Earl, R. Aguilera, D. A. Cowan, M. Ericsson, G. Gmeiner, D. Kwiatkowska, N. Kioukia-Fougia, A. Molina, J. Ruivo, J. Segura, P. Van Eenoo, N. Jan, N. Robinson, M. Saugya, Evaluation of longitudinal steroid profiles from male football players in UEFA competitions between 2008 and 2013. Drug Test. Anal. (2015) online version DOI 10.1002/dta.1851.
- [36] T. Kuuranne, M. Saugya, N. Baume, Confounding factors and genetic polymorphism in the evaluation of individual steroid profiling, Br. J. Sports Med. 48 (2014) 848–855.

Table 1. Descriptive statistics of the monitored steroids in the reference population (80 men, 16 women, aged 18–40) including the minimum, inter-quartile ranges (IQ1–IQ3), median, mean and maximum concentration values.

	Min (ng mL ⁻¹)	IQ1 (ng mL ⁻¹)	Median (ng mL ⁻¹)	Mean (ng mL ⁻¹)	IQ3 (ng mL ⁻¹)	Max (ng mL ⁻¹)
5a-Adiol	22.7	47.5	61.7	63.8	78.1	110.7
5b-Adiol	28.8	72.1	96.5	89.8	108.5	141.5
A	926	1730	2329	2382	2968	4288
DHEA	5.8	23.9	40.8	40.6	53.7	94.9
DHT	3.4	7.9	10.1	12.7	13.8	48.9
E	9.6	21.5	27.0	27.6	33.2	52.9
Etio	853	1395	1925	1847	2231	3618
T	9.4	25.3	33.1	33.6	41.3	58.9
5a-Adiol/5b-Adiol	0.2	0.6	0.7	0.8	0.9	2.2
5a-Adiol/E	0.7	1.8	2.3	2.5	2.9	5.2
A/Etio	0.4	1.0	1.2	1.3	1.6	3.0
A/T	26.9	56.7	66.0	78.1	95.4	219.8
T/E	0.9	1.1	1.2	1.2	1.3	1.9

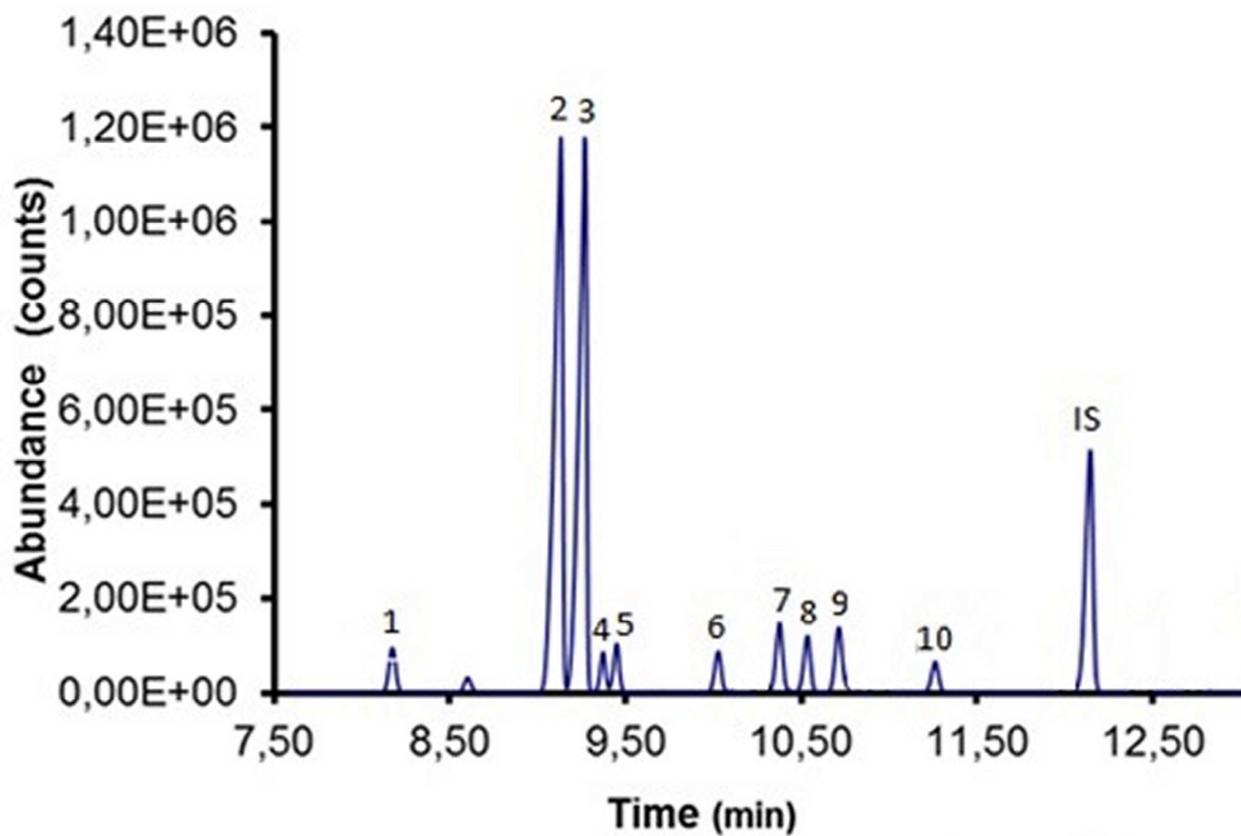


Figure 1. GC-MS chromatogram snapshot. Coded target analytes are: 1) 5β -androstane-3,17-dione; 2) androsterone; 3) etiocholanolone; 4) 5α -androstane- $3\alpha,17\beta$ -diol; 5) 5β -androstane- $3\alpha,17\beta$ -diol; 6) dehydroepiandrosterone; 7) epitestosterone; 8) 5α -dihydrotestosterone; 9) 4-androsten-3,17-dione; 10) testosterone and testosterone-d₃; IS) 17α -methyltestosterone.

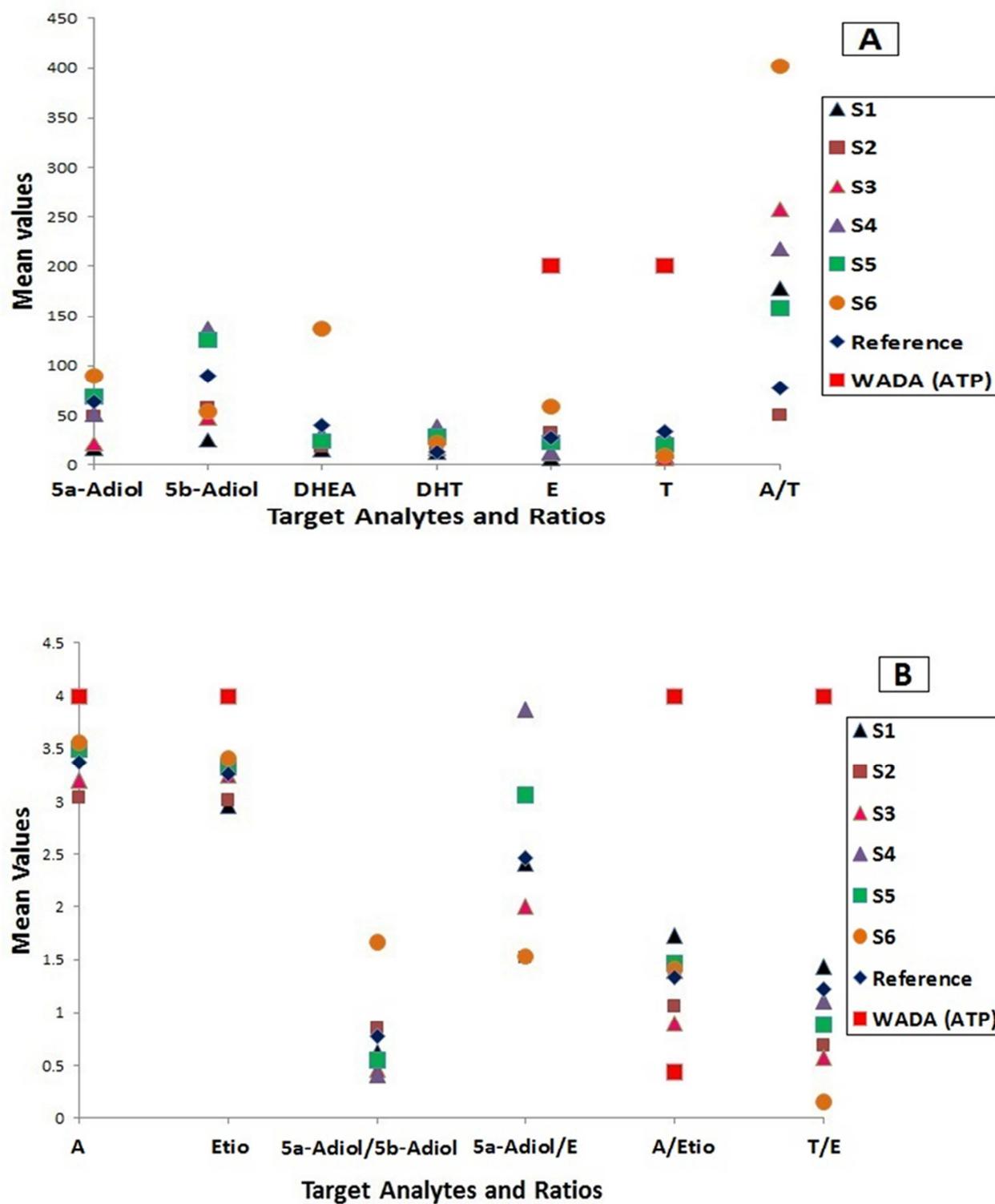


Figure 2. Mean values relevant to 8 target analytes concentration levels (ng mL^{-1}) plus 5 urinary ratios for reference population and volunteers S1-S6. WADA reference limits (when available) are indicated by red squares.

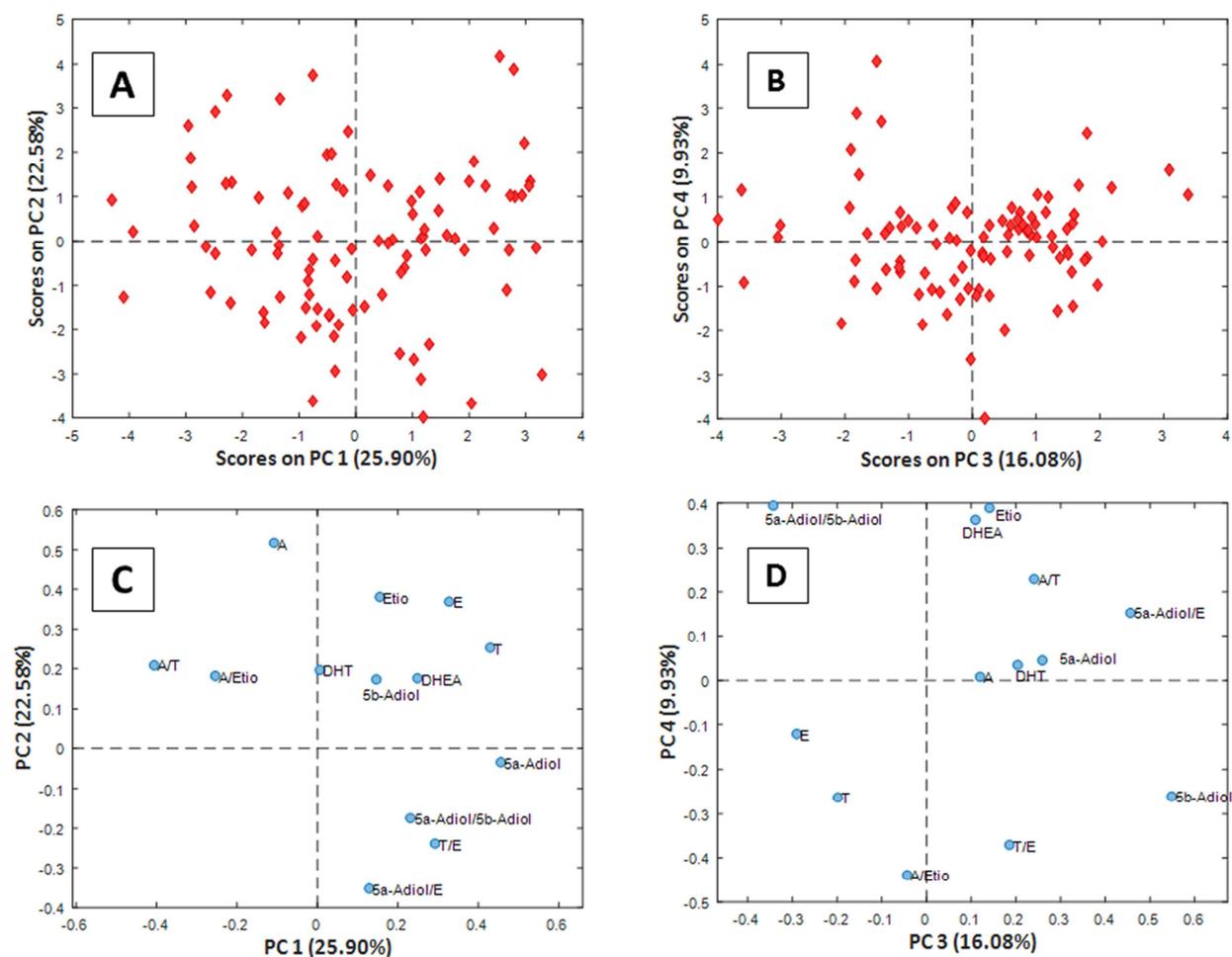


Figure 3. (A) Scores plots relevant to PC1 (V% = 25.90%) vs. PC2 (V% = 22.58%) and (B) PC3 (V% = 16.08%) vs. PC4 (V% = 9.93%); (C) loading plots relevant to PC1 vs. PC2 and (D) PC3 vs. PC4.

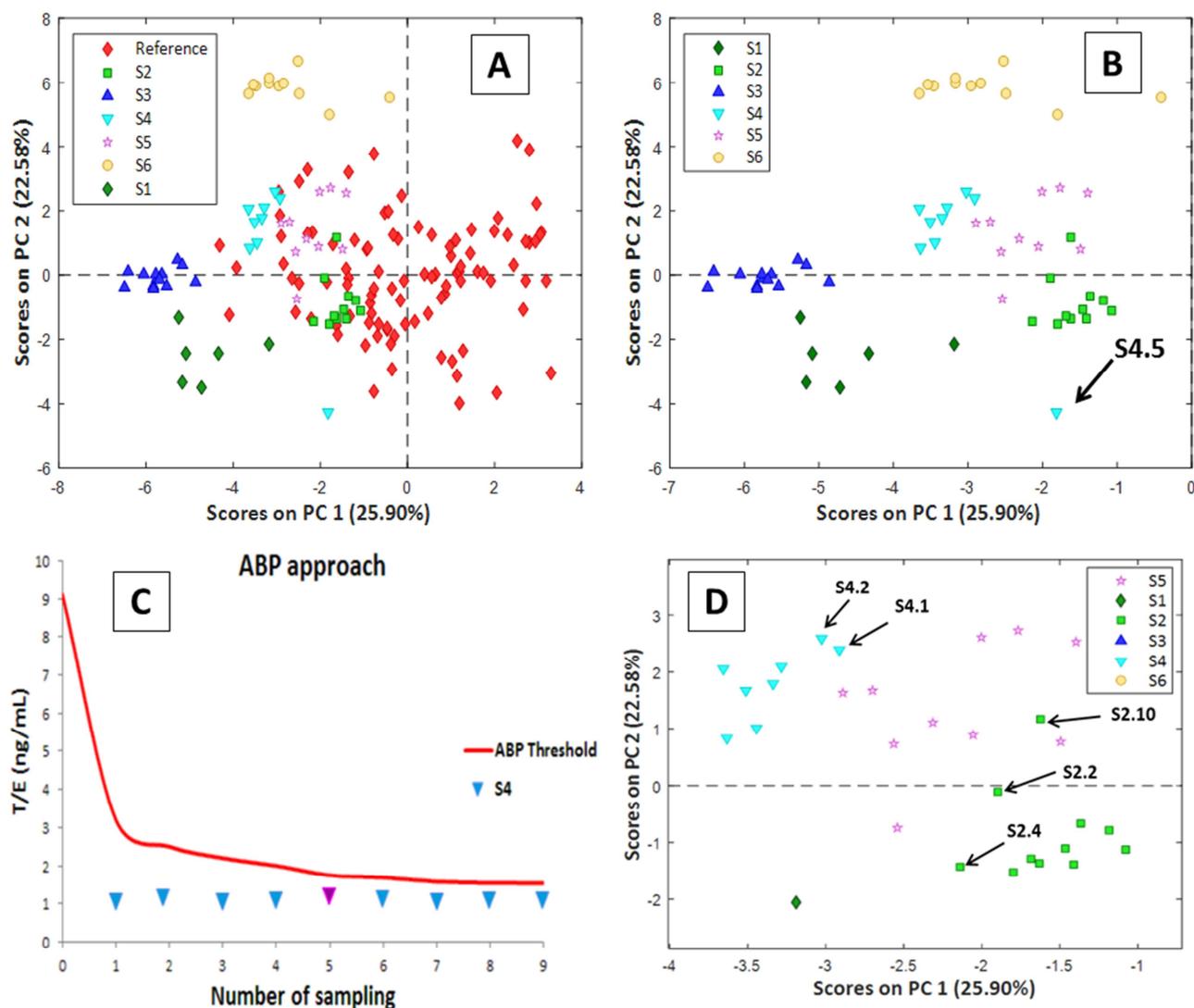


Figure 4. PC1 vs PC2 Scores Plots relevant to the time-monitored volunteers S1-S6 in the presence (A) and absence (B) of the reference population, whose urine samples are represented by red diamonds. In particular, green diamonds stand for S1, green squares for S2, blue triangles for S3, light-blue inverted triangles for S4, purple stars for S5 and yellow circles for S6. Anomalous sample S4.5 is evidenced by an arrow. (C) ABP model representing S4 samples (light blue inverted triangles) and the Bayesian threshold (red line). (D) PC1 vs PC2 Scores plot. Tested samples relative to subjects S2 (S2.2, S2.4, S2.10) and S4 (S4.1, S4.2) are indicated by light blue inverted triangles and green squares, respectively. Purple stars represents urine samples belonging to subject S5.

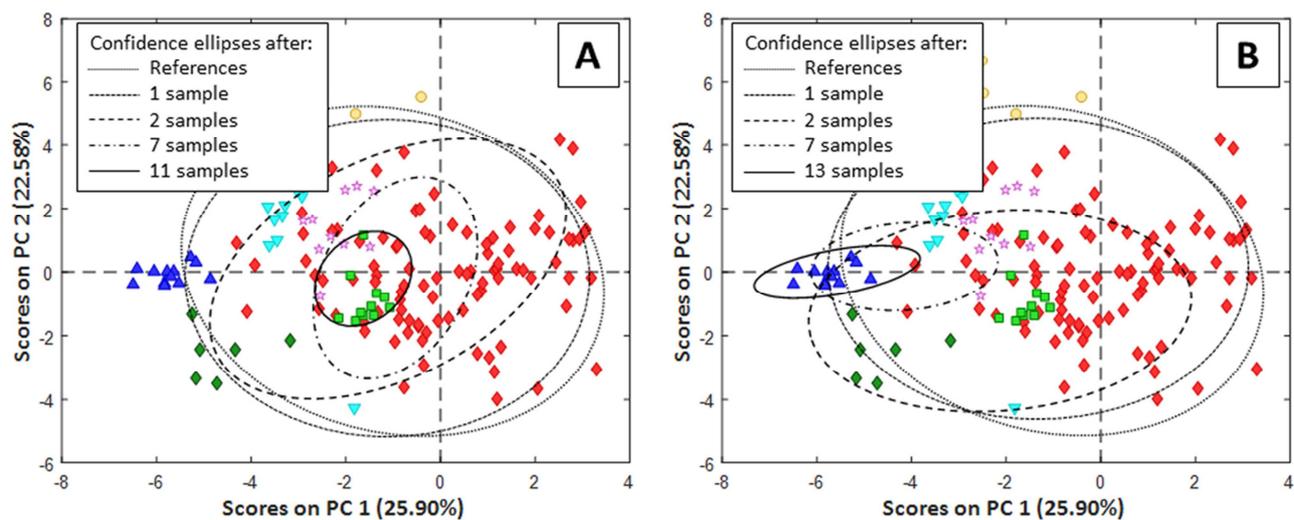


Figure 5. PC1 vs PC2 Scores plot and confidence ellipses (95% significance level). Urine samples relative to subjects S2 are indicated by green squares (A), while subjects S3 is represented by blue triangles (B). Confidence ellipses are distinguished by various dashed and continue lines; such ellipses were initially calculated on the reference population, then they were evaluated after the introduction of the first, the second, the seventh and the last urine samples from S2 and S3 time-monitored volunteers.

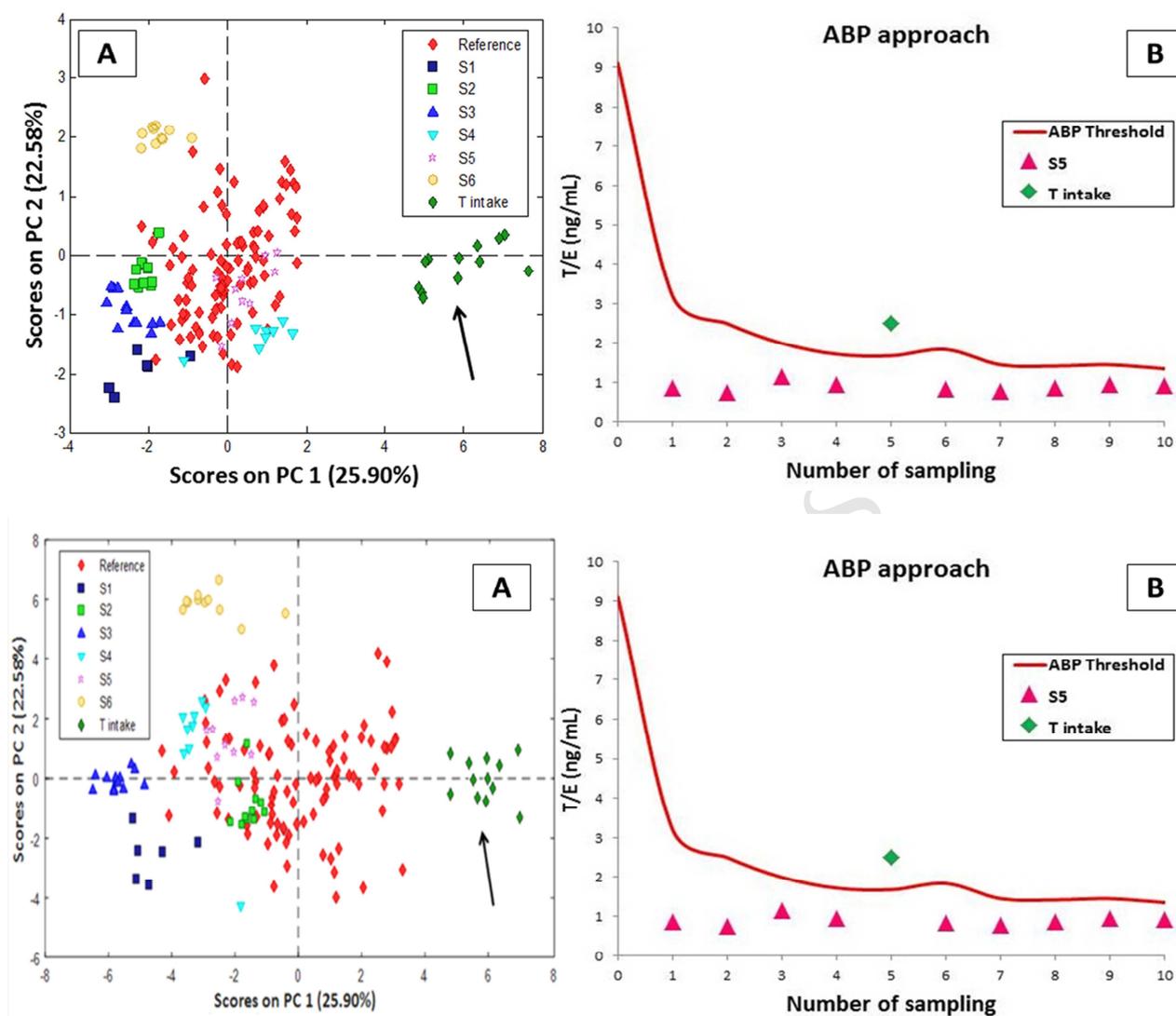


Figure 6. (A) PC1 vs PC2 Scores plot. Urine samples relative to subjects S5 are indicated by purple stars. Green diamonds (evidenced by an arrow) represent urine samples collected from Androderm® and Testovis® users. (B) ABP model representing S5 samples (purple triangles), Androderm® user sample (green diamond) and the Bayesian threshold (red line).

Detection of endogenous anabolic androgenic steroids misuse is made possible by multivariate statistical approach.

Principal Component Analysis and Hotelling's T^2 tests techniques are used to recognize anomalous values within the athletes' physiological urinary steroidal profile.

Suspect cases of urine replacement can be detected, too.

The present proof-of-concept approach might corroborate the conclusions of the Athlete Biological Passport – Steroidal Module that was drafted in 2004 by the World Anti-Doping Agency.

.