Analytica Chimica Acta 1267 (2023) 341389

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

# A new multimodal paradigm for biomarkers longitudinal monitoring: a clinical application to women steroid profiles in urine and blood

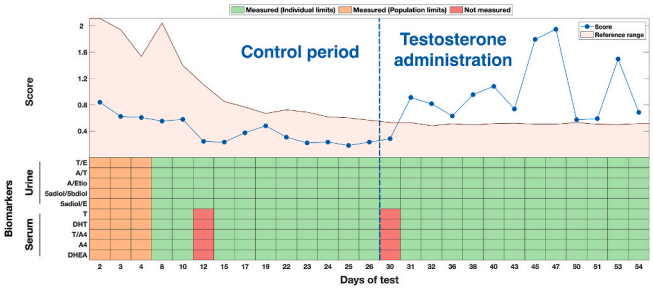
Miguel de Figueiredo [a,b], Jonas Saugy [c], Martial Saugy [c], Raphaël Faiss [c], Olivier Salamin [c,d], Raul Nicoli [d], Tiia Kuuranne [d], Serge Rudaz [a,b], Francesco Botrè [c], Julien Boccard [a,b,*]

[a] School of Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland
[b] Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Geneva, Switzerland
[c] Center of Research and Expertise in Anti-Doping Sciences, Institute of Sport Sciences, University of Lausanne, Lausanne, Switzerland
[d] Swiss Laboratory for Doping Analyses, University Center of Legal Medicine, Lausanne and Geneva, Lausanne University, Hospital and University of Lausanne, Switzerland

## HIGHLIGHTS

- A score-based approach simultaneously monitoring multiple biomarkers is proposed.
- Both between and within-variability are considered to estimate reference ranges.
- Multimodal monitoring provides a more complete evaluation of individual profiles.
- Multimodal approaches are expected to improve sensitivity in anti-doping detection.

## GRAPHICAL ABSTRACT




## ARTICLE INFO

*Keywords:*
Anti-Doping
Steroids
Blood
Urine
Multimodal
Longitudinal monitoring

## ABSTRACT

*Background:* Most current state-of-the-art strategies to generate individual adaptive reference ranges are designed to monitor one clinical parameter at a time. An innovative methodology is proposed for the simultaneous longitudinal monitoring of multiple biomarkers. The estimation of individual thresholds is performed by applying a Bayesian modeling strategy to a multivariate score integrating several biomarkers (compound concentration and/or ratio). This multimodal monitoring was applied to data from a clinical study involving 14 female volunteers with normal menstrual cycles receiving testosterone via transdermal route, as to test its ability to detect testosterone administration. The study samples consisted of urine and blood collected during 4 weeks of a control phase and 4 weeks with a daily testosterone gel application.

*Results:* Integrating multiple biomarkers improved the detection of testosterone gel administration with substantially higher sensitivity compared with the distinct follow-up of each biomarker, when applied to selected urine and serum steroid biomarkers, as well as the combination of both. Among the 175 known positive samples, 38% were identified by the multimodal approach using urine biomarkers, 79% using serum biomarkers and 83% by combining biomarkers from both biological matrices, whereas 10%, 67% and 64% were respectively detected using standard unimodal monitoring.

* Corresponding author. Institute of Pharmaceutical Science of Western Switzerland, University of Geneva, Rue Michel Servet 1, 1211, Geneva 4, Switzerland.
*E-mail address:* julien.boccard@unige.ch (J. Boccard).

https://doi.org/10.1016/j.aca.2023.341389
Received 21 March 2023; Received in revised form 12 May 2023; Accepted 16 May 2023
Available online 17 May 2023

*Significance and novelty:* The detection of abnormal patterns can be improved using multimodal approaches. The combination of urine and serum biomarkers reduced the overall number of false-negatives, thus evidencing promising complementarity between urine and blood sampling for doping control, as highlighted in the case of the use of transdermal testosterone preparations. The generation in a multimodal setting of adaptive and personalized reference ranges opens up new opportunities in clinical and anti-doping profiling. The integration of multiple parameters in a longitudinal monitoring is expected to provide a more complete evaluation of individual profiles generating actionable intelligence to further guide sample collection, analysis protocols and decision-making in clinics and anti-doping.

## 1. Introduction

One of the most challenging aspects of anti-doping science is the detection of doping using substances of exogenous origin that are also naturally present in biological fluids. Today, the longitudinal monitoring of specific biomarkers is an established approach to reveal direct or indirect effects of prohibited practices in sport. In the anti-doping field, this strategy is known as the Athlete Biological Passport (ABP), which is implemented and maintained by the World Anti-Doping Agency (WADA) in the Anti-Doping Administration and Management System (ADAMS) [1,2]. The ABP objective is not to directly target specific prohibited substances or methods, but to efficiently guide investigations and further testing of athletes, and in some cases declare anti-doping rules violations on the basis of specific patterns exhibited by the monitored biomarkers [3]. The ABP employs a Bayesian framework to grasp the dynamic nature of repeated measurements and generate individual reference ranges within which a new measure is expected to take values assuming normal physiological conditions [4].

Athletes urinary steroid profiles are investigated by longitudinally monitoring selected endogenous anabolic androgenic steroids (EAAS) composing the steroid module, as defined by the WADA [5]. Although EAAS doping detection has greatly improved using indirect pieces of evidence through the longitudinal monitoring of biomarkers impacted by fraudulent practices, it faces numerous challenges related to the nature of the urinary matrix, the employed analytical methods, as well as profile alterations due to endogenous and exogenous confounding factors [6,7]. Female steroid profiles were also shown to be affected by the menstrual cycle, pregnancy or hormonal contraception [8–10]. These reasons have led the anti-doping community to investigate new and complementary approaches for steroid profiling involving selection of sample matrices and analytical aspects. Blood being representative of the metabolic state of an individual, determining the serum steroid profile has shown greater sensitivity compared with urine [11–14].

Notwithstanding the ever-growing innovations related to the nature of the analyzed biological matrices, sample collection techniques and development of analytical methods, the statistical analysis implemented in the ABP remains somewhat limited by the monitoring of one biomarker at a time. The main disadvantage of this approach is the separated assessment of each biomarker, disregarding potential relationship between multiple biomarkers within a single profile and failing to be representative of the status of an individual. To tackle these limitations, a new paradigm is proposed for the generation of adaptive reference ranges based on the simultaneous longitudinal monitoring of multiple biomarkers through a multivariate score. This approach was applied to the clinical study data of Salamin et al. involving women volunteers with normal and regular menstrual cycles receiving testosterone via gel administration [15]. Performance of a multimodal detection of testosterone gel application was evaluated in comparison with the results obtained with a standard unimodal approach, such as the one implemented in the current steroidal module of the ABP. Although several parameters are considered by athlete passport management units to determine if a profile is suspicious or not, in practice the integration of multiple steroid biomarkers mostly relies on human expert knowledge. The proposed methodology objectively integrates multiple biomarkers to flag atypical profiles within a statistical framework. These tools were applied to women steroid biomarker panels in urine, blood and in a combination of both.

## 2. Materials and methods

### 2.1. Generating adaptive reference ranges

A reference range defines the interval in which a biomarker can be expected to take values under what is considered to be normal physiological conditions. The most standard reference ranges are based on a reference population [16]. In this setting, reference ranges fail to describe the expected variability of a particular individual and mix within and between-subject variability. The need to distinguish the within-subject variability led to the development of several approaches for the detection of abnormal physiological variations. Some approaches estimate empirical means and variances based solely on individual past measurements, disregarding population-based information, whereas others incorporate an overall within-subject variability fixed for all subjects [17]. However, these approaches do not integrate both between and within-subject variability. To circumvent this limitation, the Bayesian framework was used to combine general population and individual measurements by sequentially estimating adaptive reference ranges for each new measure from a given subject. The seminal method using a Bayesian framework for detecting abnormal values in anti-doping was developed by Sottas et al. and has been extensively used within the anti-doping program [18–21]. Nevertheless, as stated by Roshan et al., the above-mentioned approaches are limited by the assumption that an equal within-subject variability is considered for all subjects [22]. The Bayesian approach proposed by Roshan et al., and briefly described below, has the advantage of considering different within-subject variabilities for the longitudinal monitoring of one biomarker at a time, which is defined here as the unimodal approach.

Let $I$ be the total number of individuals sampled from a population. For each individual $i$, $n_i$ measurements are available, and $y_{ij}$ defines the jth measurement of the i$^{th}$ individual. It is assumed that clinical parameters with unknown mean $\mu_i$ and variance $\sigma_i^2$ follow a normal, or log-normal, distribution (Eq. (1)).

$$y_{ij}|\mu_i, \sigma_i^2 \sim N(\mu_i, \sigma_i^2), \text{for } i = 1, \ldots, I \text{ and } j = 1, \ldots, n_i \quad (1)$$

Individuals can have different means $\mu_i$, but it is assumed that they share a common distribution with mean $\mu$ and variance $\tau^2$, from which they are sampled. Individual dynamic reference ranges can be estimated if there are at least 2 past measurements. If not, it is reasonable to assume that $\mu_i$ follows a distribution with parameters $\mu$ and $\tau^2$, *i.e.* population-based values. For each new measurement of individual $i$, the estimation of the unknown parameters $\mu_i$ and $\sigma_i^2$ provides information on the central tendency of a person's mean and dispersion around this value, characterizing the within-subject variability. Roshan et al. proposed estimating the values a new measurement $y_{i(n_i+1)}$ can take using a Gibbs sampler. First, an estimate of the general population mean $\mu$ is drawn from the distribution in Eq. (2).

$$p(\mu|\tau^2,\boldsymbol{\sigma}^2,\boldsymbol{\mu},Y) \sim N\left(\frac{\frac{\sum_{i=1}^{I}\mu_i}{\tau^2}}{\frac{1}{\upsilon^2}+\frac{I}{\tau^2}},\frac{1}{\frac{1}{\upsilon^2}+\frac{I}{\tau^2}}\right) \quad (2)$$

where $\upsilon^2$ is defined as a constant set to cover a wide range of possible values of μ. When $I$ increases, the central tendency parameter of Eq. (2) tends to the mean of all individual means $\sum_{i=1}^{I}\mu_i/I$, and the variance parameter is mostly influenced by the ratio $\tau^2/I$. Hence, the larger the number of individuals $I$, the greater the confidence on the estimation of the central tendency parameter, which is injected in Eq. (3) to estimate the general population variance $\tau^2$ using an Inverse Gamma (IG) distribution.

$$p(\tau^2|\mu,\boldsymbol{\sigma}^2,\boldsymbol{\mu},Y) \sim IG\left(\alpha_2+\frac{I}{2},\beta_2+\frac{1}{2}\sum_{i=1}^{I}(\mu_i-\mu)^2\right) \quad (3)$$

where $\alpha_2$ and $\beta_2$ are defined as vague priors for the between-subjects variability, and correspond to the shape and scale parameters of the distribution, respectively. The larger the number of individuals $I$, the higher the shape parameter, resulting in a taller probability density function (PDF) with a thinner tail. Then, an estimate of the individual variance $\sigma_i^2$ is sampled from Eq. (4) using an IG distribution.

$$p(\sigma_i^2|\mu,\tau^2,\boldsymbol{\sigma}^2_{-i},\boldsymbol{\mu},Y) \sim IG\left(\alpha_1+\frac{n_i}{2},\beta_1+\frac{1}{2}\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2\right) \quad (4)$$

where $\alpha_1$ and $\beta_1$ are defined as vague priors for the within-subject variability. The scale parameter is mostly dominated by the number of individual past measurements $n_i$, and the shape parameter by the deviation of each measurement $y_{ij}$ to the individual mean $\mu_i$. The $\sigma_i^2$ value drawn from Eq. (4) is injected in Eq. (5) to estimate the individual mean $\mu_i$.

$$p(\mu_i|\tau^2,\boldsymbol{\sigma}^2,\boldsymbol{\mu}_{-i},Y) \sim N\left(\frac{\frac{\mu}{\tau^2}+\frac{\sum_{j=1}^{n_i}y_{ij}}{\sigma_i^2}}{\frac{1}{\tau^2}+\frac{n_i}{\sigma_i^2}},\frac{1}{\frac{1}{\tau^2}+\frac{n_i}{\sigma_i^2}}\right) \quad (5)$$

When $n_i$ is zero, the central tendency and variance parameters are dominated by the general population. When $n_i$ increases, these two parameters tend to be defined by the mean of all individual measurements $\sum_{j=1}^{n_i}y_{ij}/n_i$ and the ratio $\sigma_i^2/n_i$, respectively. The values drawn from Eqs. (4) and (5) are then injected in Eq. (1) to estimate a new measurement $y_{i(n_i+1)}$. The Gibbs sampler iterates multiple times, *e.g.* 10′000, leading to a distribution of values that a new measurement $y_{i(n_i+1)}$ is expected to take. The individual reference range is then defined by the quantiles of the estimated values for $y_{i(n_i+1)}$, based on a given alpha risk.

Saulière et al. proposed multivariate extensions of individual scores using a single rule decision to avoid multiple testing issues [23]. The suggested methods are based on individual past measurements, but do not integrate population-based information. Moreover, a minimum number of measurements is required to start estimating individual reference ranges, which increases with the number of monitored biomarkers. An extension of the method proposed by Roshan et al. to a multivariate setting is thus proposed based on a transformation of the variable to be monitored. Whereas the concentration of a single biomarker is followed in the unimodal strategy, the multimodal approach integrates multiple biomarkers through a multivariate criterion, referred to as a *score* hereafter. It summarizes the information contained in these biomarkers using a continuous metric, which can take many forms such as distance or similarity measures. It is assumed that the chosen metric follows a normal or log-normal distribution. A score is measured between pairs of measurements, meaning that every new measurement $\boldsymbol{y}_{i(n_i+1)}$ is compared with a reference measurement $\boldsymbol{y}_{ref}$. This reference measurement contains the expected biomarker concentrations with respect to the past $n_i$ measurements of individual $i$ calculated as $\sum_{j=1}^{n_i}\boldsymbol{y}_{ij}/\boldsymbol{n}_i$ for each biomarker. Note that for the first two points, $\boldsymbol{y}_{ref}$ is calculated based on the general population as $\sum_{i=1}^{I}\mu_i/I$, where $\mu_i$ is the mean of individual $i$, and $I$ the total number of individuals. A Euclidean distance-based metric was chosen to express the difference between $\boldsymbol{y}_{i(n_i+1)}$ and $\boldsymbol{y}_{ref}$, whose score is calculated as in Eq. (6). Euclidean distances have the advantage of being simple and largely used in machine learning methods to assess dissimilarities between pairs of observations. Other distance or similarity measures, possibly more complex, could be investigated, but for sake of simplicity, a straightforward metric was favored.

$$score_{i(n_i+1)} = \left(\left(\boldsymbol{y}_{ref}-\boldsymbol{y}_{i(n_i+1)}\right)\bullet\boldsymbol{C}^{-1}\bullet\left(\boldsymbol{y}_{ref}-\boldsymbol{y}_{i(n_i+1)}\right)^{\mathrm{T}}\right)^{1/2} \quad (6)$$

where $(\boldsymbol{y}_{ref}-\boldsymbol{y}_{i(n_i+1)})$ is the element-wise difference between the biomarker concentrations in row vectors $\boldsymbol{y}_{i(n_i+1)}$ and $\boldsymbol{y}_{ref}$, $\boldsymbol{C}$ is a p-by-p square matrix, where p is the number of biomarkers. If $\boldsymbol{C}$ is an identity matrix with zeros everywhere, except for the diagonal elements, which are ones, then standard Euclidean distance is obtained. Replacing the ones in the diagonal by positive scale factors for each dimension, *e.g.* the standard deviation of the past measurements for each biomarker, yields a scaled Euclidean distance. If $\boldsymbol{C}$ is a covariance matrix of the past measurements and assumed to be invertible, then relationships between biomarkers can be considered in the calculations. This score monitors deviations from the expected values. If the deviations of a new measurement $\boldsymbol{y}_{i(n_i+1)}$ are within the expected reference range, then normal physiological conditions are assumed, but if it falls outside the expected reference range, an alert is triggered for a possible abnormal variation.

### 2.2. Experimental data

Urine and blood were collected from 14 women volunteers with normal and regular menstrual cycles over two periods of 28 days [15]. Although the original work contained samples collected during 3 phases, only the first two were kept in this work. For each volunteer and according to the clinical study design in Fig. 1, 15 samples were collected without testosterone administration during the 1st cycle, serving thus as a control period with natural variations of the monitored biomarkers. During the 2nd cycle, testosterone was daily administered via gel, then urine and blood samples were collected at similar time points. Given that not all 420 scheduled samples could be collected, the final dataset contained a total of 391 samples, including 216 known negative samples and 175 known positive samples. Samples are labeled as known positives when it is confirmed by the experimental design that testosterone was administered, except for day 29 given that collection was made before the administration. Serum steroid concentrations were determined with a validated method using ultra-high-performance liquid chromatography coupled to tandem mass-spectrometry. Urine steroid concentrations were determined with a validated method using gas chromatography coupled to mass-spectrometry. Further information about the analytical procedures, eligibility and consent of the study cohort, as well as clinical trial design and approval can be found in Salamin et al. [15].

### 2.3. Comparison framework

The urinary and the serum steroid profiles determined by Salamin et al. were used to compare the unimodal and multimodal approaches. The 5 ratios of biomarkers included in ADAMS' ABP steroidal module were used to build the longitudinal profiles from the urinary steroid measurements, namely testosterone to epitestosterone (T/E), androsterone to testosterone (A/T), androsterone to etiocholanolone (A/Etio),

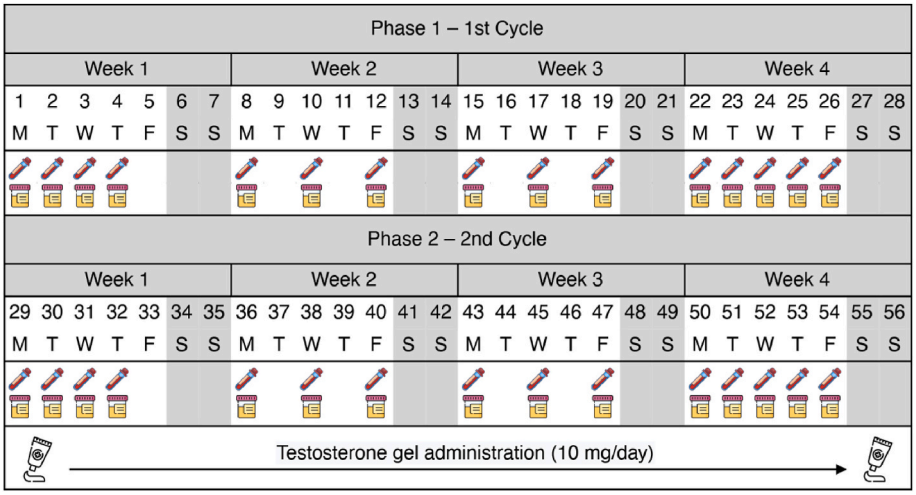



**Fig. 1.** Design of the clinical study conducted by Salamin et al. involving women volunteers with normal menstrual cycles receiving testosterone via gel administration [15]. In each cycle, the weeks and days where samples were collected are marked with blood and urine symbols.

5α-androstane-3α,17β-diol to 5β-androstane-3α,17β-diol (5aAdiol/5bAdiol) and 5α-androstane-3α,17β-diol to epitestosterone (5aAdiol/E). The serum steroid profile was evaluated on the basis of 4 biomarkers and 1 ratio, namely testosterone (T), androstenedione (A4), dihydrotestosterone (DHT), dehydroepiandrosterone (DHEA) and testosterone to androstenedione (T/A4). The longitudinal monitoring was then investigated for its ability to detect exogenous testosterone administration by applying either unimodal or multimodal monitoring of the biomarkers included in: (1) the urinary steroid profile; (2) the serum steroid profile; (3) the combination of both urinary and serum steroid profiles. To consider that multiple tests are performed in the unimodal setting, a Bonferroni correction was applied on the alpha risk since a single biomarker outside the reference range was defined as a sufficient condition to declare a sample as atypical. For each sample of each volunteer, the reference ranges were estimated using the known negative samples of the whole cohort for the first 3 time points, then using individual past measurements. Because a given sample is not involved in the estimation of its own reference range, it is used as an independent test. The predictions for all samples are then assessed using true and false positive rates (TPR and FPR). TPR are known positive samples, whose outcome was outside the reference ranges during the testosterone administration phase. FPR are known negative samples, whose outcome was outside the reference ranges during the control phase. The false negative rate (FNR) constitutes another important parameter to be considered for detection purposes. In the context of such a controlled clinical study, it can easily be derived as the complement of the true positive rate TPR (FNR + TPR = 100%).

The processing of the longitudinal profiles was done at an alpha risk of 0.001 (0.1%). Biomarkers were scaled to unit length using past measurements and a scaled Euclidean distance was used as score (Eq. (6)), where the diagonal of the scaling matrix $\boldsymbol{C}$ contains the standard deviation of the biomarkers for the past measurements.

## 3. Results and discussion

True and false positive rates obtained by processing all volunteers' longitudinal data from the two first phases are shown in Fig. 2. Considering only urine (Fig. 2A), FPR results were comparable, but the integration of all biomarkers in a multimodal longitudinal monitoring led to a greater sensitivity (38%) than the unimodal one (10%). Regarding serum (Fig. 2B), FPR results were similar to urine, but a substantially higher TPR was found for both unimodal (68%) and multimodal (79%) approaches. It is to be noted that a greater sensitivity in urine and serum was obtained with the latter. As for the integration of all biomarkers from both biological matrices (Fig. 2C), FPR results are similar with slightly higher values than previously. A substantially higher sensitivity was found for the multimodal strategy (83%) compared with unimodal one (64%).

Unimodal approaches are designed to monitor one biomarker at a time, whereas multimodal strategies can follow and integrate several biomarkers simultaneously. In this context, the multimodal monitoring offered greater sensitivity and similar selectivity compared with the unimodal approach, indicating that biomarkers should be integrated in a multivariate score. The chosen conservative approach with an alpha risk at 0.1% led to detect most of the samples collected following an administration of testosterone gel. For cases where a higher FPR may be acceptable, increasing the alpha risk would lead to higher FPR, but also higher TPR. Unimodal approaches can be used to monitor biomarkers separately but the integration of the results needs to be corrected for multiple testing issues. Consequently, when combining urine and serum

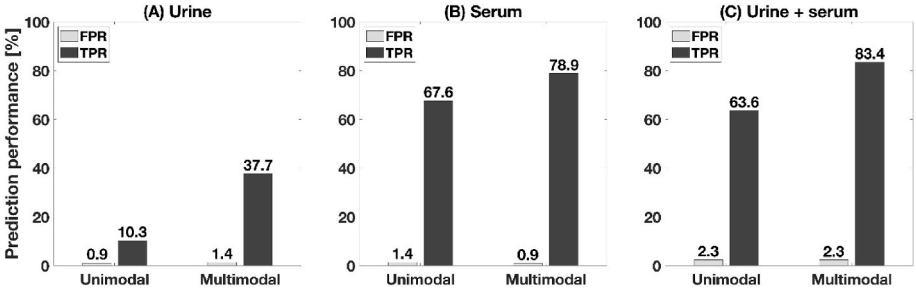



**Fig. 2.** False positive rate (FPR) and true positive rate (TPR) for the unimodal and multimodal approaches when applied to urine (A), serum (B) and both urine plus serum (C) biomarkers.

biomarkers, a lower TPR was obtained compared with serum alone (Fig. 2C). The multimodal approach avoids this issue by integrating the information on the deviations of all biomarkers in a single score. The increase in sensitivity of the multimodal approach applied to the biomarkers in urine, serum and a combination of both suggests complementarity between both matrices with respect to the detection of doping, and interpretation.

Fig. 3 shows a grayscale heatmap of the multimodal results of both urine and serum biomarkers for all volunteers during all collection days. Samples are defined as non-collected (NA), in-range or out-of-range. This heatmap shows that testosterone administration via gel could be evidenced for all volunteers, indicating that the individual multimodal monitoring is efficient and does not suffer from potential heterogeneity between the volunteers.

Fig. 4 shows the profiles of volunteer 5 for urine (Fig. 4A) and serum (Fig. 4B) biomarkers, as well as the combination of all urine and serum biomarkers (Fig. 4C). Multimodal longitudinal profiles show reference ranges as a shaded area with scores shown as circle markers for each day of test. The colored grid below shows missing values as red cells. Orange and green cells are associated with measured values for which reference ranges were based on the general population measures (orange) or on individual past measurements (green). In Fig. 4, the grid cells of the first 3 days are orange, the rest is green, except for days 12 and 30 in Fig. 4B, which are red. The urine longitudinal profile of volunteer 5 (Fig. 4A) shows that day of test 53 is the only one outside the reference range. These limits gradually decrease with the increasing number of samples, suggesting that the within-volunteer variability is being modeled. The serum longitudinal profile (Fig. 4B) shows more samples detected outside the limits, as expected from results in Fig. 2. Although the reference ranges tend to decrease with the increase of the number of measurements, they remain higher between days of test 8 and 15 with respect to the general population limits. This is the consequence of a high within-variability with respect to the past measurements of volunteer 5, and a higher uncertainty in the estimation of individual parameters. Also, two blood samples were not collected on days 12 and 30. The multimodal approach can handle missing values by replacing them by their expected value. For this reason, the scores of days 12 and 30 are zero. This conservative approach to handle missing data is paramount because real data are often incomplete. The absence of a measure can be due to the inability to collect a sample, a biomarker concentration that was not measured, or its value was below the limit of quantification. It is noteworthy that in practice, days of test 12 and 30 of the serum profile should not be included given that no biomarker was measured. However, they are kept in this example for visualization purposes and the combination of urine and serum biomarkers. The complementarity between urine and serum can be observed in Fig. 4C. Days of test 50, 51 and 54 are in-range in both urine and serum profiles separately, although having higher scores than past measurements and closer to the reference limits. The integration of all 10 biomarkers combines the deviations resulting in these points being outside the reference range.

Because of their synthetic nature, multimodal scores do not directly reflect concentrations of biomarkers involved in the longitudinal monitoring. It is, however, possible to retrieve their specific contribution as to understand the reason(s) for an outcome being outside the reference range. This property is of utmost importance when interpreting longitudinal profiles, and the associated information can be obtained by investigating the residuals. The residuals are the element-wise squared differences between a new measurement and the reference. The residuals of day 53 for each of the profiles in Fig. 4 are shown in Fig. 5. For the urine profile, the ratios 5adiol/E and T/E were mostly responsible for the sample being out-of-range. Indeed, exogenous T administration increases the T/E ratio induced by a combination of intrinsic T increase and E decrease from negative feedback. Historically, an initial threshold of 6 based on population studies was defined to reflect testosterone administration [24,25]. This threshold was then leveled to 4 by WADA in 2004 [26]. However, the most significant sensitivity increase was achieved when the T/E was implemented in the steroidal module of the ABP in 2014, further improved with the use of the 5α-diol/E [27]. In serum, as recently demonstrated by several authors, T and the T/A4 ratio are out of the most promising and sensitive markers to be included in an individualized and longitudinal anti-doping approach [12–14,28]. When combining all biomarkers in both matrices, the same biomarkers were evidenced with the serum biomarkers having more weight.

The longitudinal monitoring of multiple biomarkers comes, nevertheless, with challenges. Real world data may be measured in different units or span different concentration ranges, impacting the calculation of the score. When biomarkers have different scales, higher values will tend to dominate the score. For this reason, biomarkers should be scaled to correct for those differences, and common approaches are the min-max normalization, standardization or scaling to unit length. Also, the inclusion of too many irrelevant or noisy parameters with naturally high variability may reduce the efficiency of a multimodal strategy. Moreover, the presence of many missing values can hinder its intelligence potential since a proper longitudinal monitoring can only be achieved when biomarkers are complete and regularly monitored. In the context of the clinical study presented here, a daily sampling rate was applied for each volunteer, which is not representative of the data structure usually

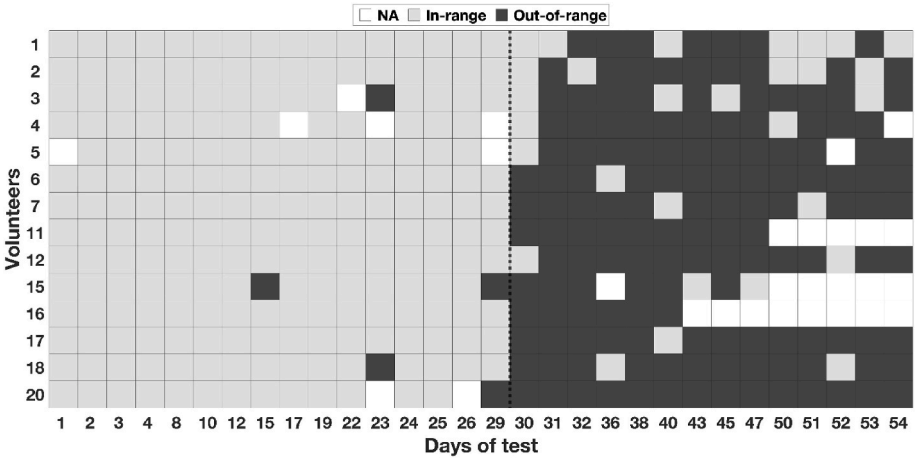



**Fig. 3.** Heatmap of the position of each collected sample with respect to the reference ranges of the multimodal longitudinal profiles. Samples that could not be collected are defined as not assigned (NA). The dotted-line between days 29 and 30 indicates the separation between known negative samples (left) and known positive samples (right).

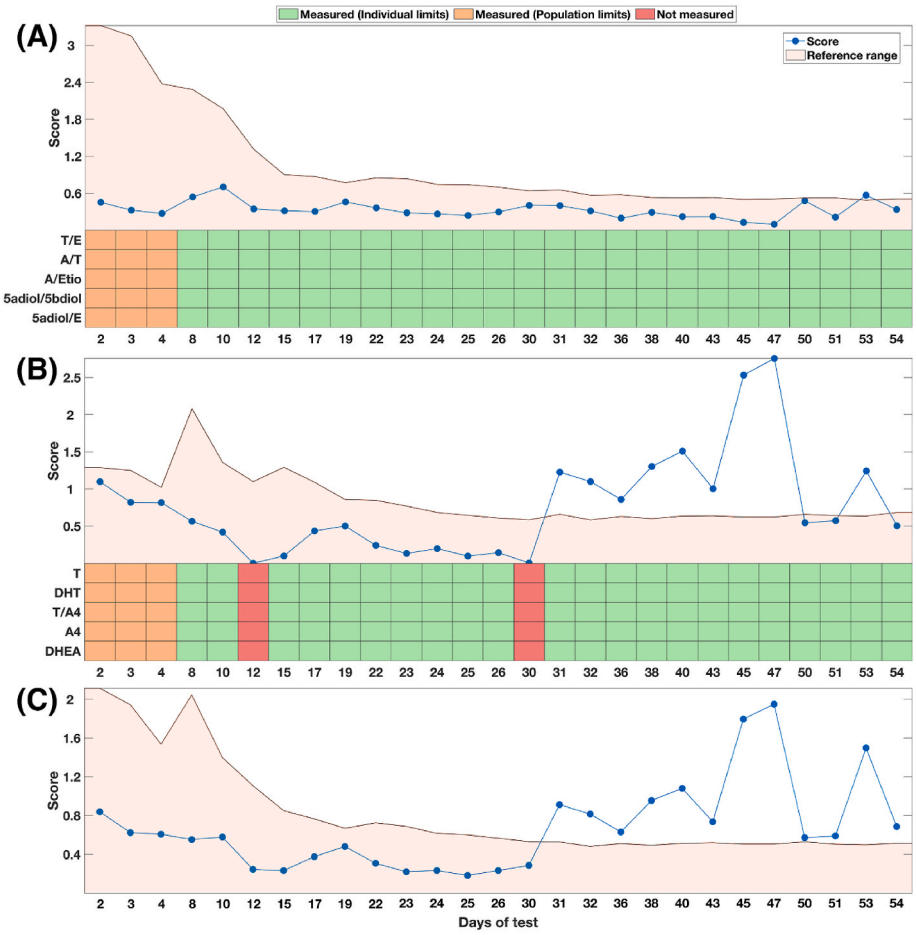



**Fig. 4.** Longitudinal profiles of volunteer 5 for urine (A) and serum (B) biomarkers, as well as a combination of all urine and serum biomarkers (C).

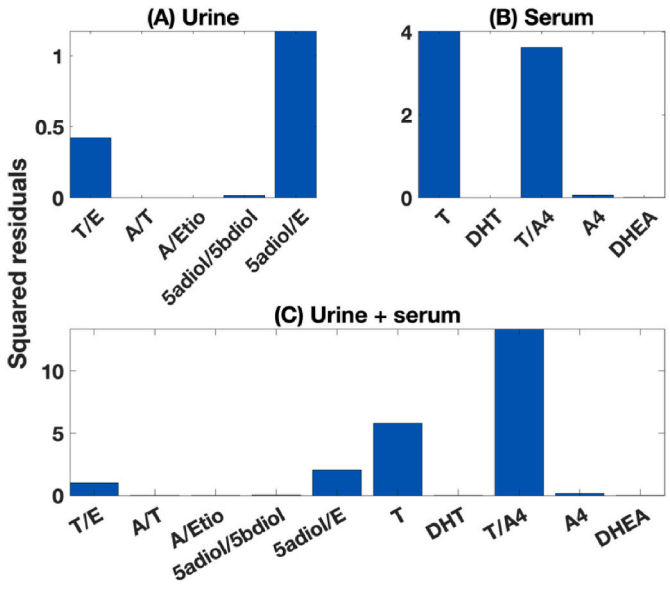



**Fig. 5.** Residuals as the element-wise squared differences between day of test 53 and the associated reference values for urine (A) and serum (B) biomarkers, as well as a combination of all urine and serum biomarkers (C).

available within the ABP. In fact, the time between two sample collections for a given athlete may span days, months or even years. In anti-doping, higher sampling rates also increase the burden for testing organizations to regularly collect different types of sample matrices. This is, nevertheless, essential to build representative athlete longitudinal profiles because large intervals between successive tests may artificially increase within-athlete variability, thus generating wider reference ranges, and higher FNR. Contemporaneity of sample collection was shown in this work to be paramount to detect abnormal variations in urinary and blood steroid profiles, which also constitutes an inherent constraint of the ABP steroidal module and its routine implementation.

## 4. Conclusions

The generation of adaptive and personalized reference ranges in a multimodal setting opens up new opportunities in anti-doping profiling. In this study, a substantial gain in sensitivity was observed for the multimodal approach compared with the unimodal strategy. Tools able to integrate information from multiple biomarkers are thus needed to provide a more complete picture of the metabolic condition of an individual. Furthermore, multimodal approaches yield a more easily interpretable graphical representation of longitudinal profiles in one chart, instead of one for each biomarker, while avoiding multiple testing issues. Determining which biomarkers, preprocessing methods and scores are the most appropriate and robust to build longitudinal profiles would further improve the efficiency of the strategy. This does not imply, however, that multimodal methods are meant to replace unimodal ones. Preserving the initial measurement units in univariate profiles is indeed of great interpretation value, and this makes both approaches complementary. Addressing the implementation challenges of unimodal and multimodal methodologies is key to generate actionable intelligence to further guide sample collection, analysis protocols and decision-making

in anti-doping. Beyond anti-doping sciences, longitudinal monitoring of biomarkers can provide information about the health status of individuals. This may be particularly beneficial to guide medical decisions for diagnosis, progress monitoring and treatment evaluation of pathologies. Such a strategy is expected to improve patient care by taking an additional step towards personalized medicine, and will be investigated within a clinical setting. Future work will also incorporate covariates (*e.g.* age, gender, time of day, laboratories) in the estimation procedure to help interpreting the results and control for potential confounding factors, an important aspect to interpret longitudinal profiles.

## Funding

This work was supported by the Athletics Integrity Unit, Monaco.

## CRediT authorship contribution statement

**Miguel de Figueiredo:** Conceptualization, Formal analysis, Methodology, Project administration, Software, Visualization, Writing – original draft, Writing – review & editing. **Jonas Saugy:** Conceptualization, Formal analysis, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Martial Saugy:** Conceptualization, Funding acquisition, Methodology, Project administration, Writing – review & editing. **Raphaël Faiss:** Conceptualization, Methodology, Project administration, Writing – review & editing. **Olivier Salamin:** Conceptualization, Data curation, Methodology, Project administration, Writing – review & editing. **Raul Nicoli:** Conceptualization, Methodology, Project administration, Writing – review & editing. **Tiia Kuuranne:** Conceptualization, Methodology, Project administration, Writing – review & editing. **Serge Rudaz:** Conceptualization, Funding acquisition, Methodology, Project administration, Writing – review & editing. **Francesco Botrè:** Conceptualization, Funding acquisition, Methodology, Project administration, Writing – review & editing. **Julien Boccard:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

## Declaration of competing interest

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.

## Data availability

The authors do not have permission to share data.

## References

[1] Wada, Athlete Biological Passport Operating Guidelines, World Anti-Doping Agency, Montreal, Canada, 2021. https://www.wada-ama.org/sites/default/files/resources/files/guidelines_abp_v8_final.pdf. (Accessed 21 November 2022).
[2] N. Robinson, M. Saugy, A. Vernec, P.-E. Sottas, The athlete biological passport: an effective tool in the fight against doping, Clin. Chem. 57 (2011) 830–832, https://doi.org/10.1373/clinchem.2011.162107.
[3] P.-E. Sottas, N. Robinson, O. Rabin, M. Saugy, The athlete biological passport, Clin. Chem. 57 (2011) 969–976, https://doi.org/10.1373/clinchem.2011.162271.
[4] M. Saugy, C. Lundby, N. Robinson, Monitoring of biological markers indicative of doping: the athlete biological passport, Br. J. Sports Med. 48 (2014) 827–832, https://doi.org/10.1136/bjsports-2014-093512.
[5] WADA, Technical document TD2021EAAS, Measurement and Reporting of Endogenous Anabolic Androgenic Steroid (EAAS) Markers of the Urinary Steroid Profile, World Anti-Doping Agency, Montreal, Canada, 2021. https://www.wada-ama.org/sites/default/files/2022-01/td2021eaas_final_eng_v_2.0.pdf. (Accessed 13 October 2022). accessed.
[6] F. Ponzetto, N. Baume, C. Schweizer, M. Saugy, T. Kuuranne, Steroidal module of the athlete biological passport, Curr. Opin. Endocr. Metab. Res. 9 (2019) 14–21, https://doi.org/10.1016/j.coemr.2019.05.006.
[7] T. Kuuranne, M. Saugy, N. Baume, Confounding factors and genetic polymorphism in the evaluation of individual steroid profiling, Br. J. Sports Med. 48 (2014) 848–855, https://doi.org/10.1136/bjsports-2014-093510.
[8] J.E. Mullen, J.-O. Thörngren, J.J. Schulze, M. Ericsson, N. Gårevik, M. Lehtihet, L. Ekström, Urinary steroid profile in females – the impact of menstrual cycle and emergency contraceptives, Drug Test. Anal. 9 (2017) 1034–1042, https://doi.org/10.1002/dta.2121.
[9] J. Mullen, Y. Gadot, E. Eklund, A. Andersson, J.J. Schulze, M. Ericsson, A. Lindén Hirschberg, A. Rane, L. Ekström, Pregnancy greatly affects the steroidal module of the Athlete Biological Passport, Drug Test. Anal. 10 (2018) 1070–1075, https://doi.org/10.1002/dta.2361.
[10] J.J. Schulze, J.E. Mullen, E. Berglund Lindgren, M. Ericsson, L. Ekström, A. L. Hirschberg, The impact of genetics and hormonal contraceptives on the steroid profile in female athletes, Front. Endocrinol. 5 (2014). https://www.frontiersin.org/articles/10.3389/fendo.2014.00050. (Accessed 17 January 2023).
[11] F. Ponzetto, J. Boccard, R. Nicoli, T. Kuuranne, M. Saugy, S. Rudaz, Steroidomics for highlighting novel serum biomarkers of testosterone doping, Bioanalysis 11 (2019) 1169–1185, https://doi.org/10.4155/bio-2019-0079.
[12] F. Ponzetto, F. Mehl, J. Boccard, N. Baume, S. Rudaz, M. Saugy, R. Nicoli, 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, https://doi.org/10.1007/s00216-015-9185-1.
[13] S. Savkovic, L.P. Ly, R. Desai, J. Howa, V. Nair, D. Eichner, D.J. Handelsman, Detection of testosterone microdosing in healthy females, Drug Test. Anal. 14 (2022) 653–666, https://doi.org/10.1002/dta.3202.
[14] J. Elings Knutsson, A. Andersson, L.V. Baekken, A. Pohanka, L. Ekström, A. L. Hirschberg, Disposition of urinary and serum steroid metabolites in response to testosterone administration in healthy women, J. Clin. Endocrinol. Metab. 106 (2021) 697–707, https://doi.org/10.1210/clinem/dgaa904.
[15] O. Salamin, R. Nicoli, T. Langer, J. Boccard, C.S. Grundisch, C. Xu, S. Rudaz, T. Kuuranne, N. Pitteloud, M. Saugy, Longitudinal evaluation of multiple biomarkers for the detection of testosterone gel administration in women with normal menstrual cycle, Drug Test. Anal. 14 (2022) 833–850, https://doi.org/10.1002/dta.3040.
[16] M. Pusparum, G. Ertaylan, O. Thas, Individual reference intervals for personalised interpretation of clinical and metabolomics measurements, J. Biomed. Inf. 131 (2022), 104111, https://doi.org/10.1016/j.jbi.2022.104111.
[17] K. Sharpe, M.J. Ashenden, Y.O. Schumacher, A third generation approach to detect erythropoietin abuse in athletes, Haematologica 91 (2006) 356–363, https://doi.org/10.3324/%x.
[18] 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, Biostatistics 8 (2007) 285–296, https://doi.org/10.1093/biostatistics/kxl009.
[19] N. Robinson, P.-E. Sottas, P. Mangin, M. Saugy, Bayesian detection of abnormal hematological values to introduce a no-start rule for heterogeneous populations of athletes, Haematologica 92 (2007) 1143–1144, https://doi.org/10.3324/haematol.11182.
[20] T. Pottgiesser, P.-E. Sottas, T. Echteler, N. Robinson, M. Umhau, Y.O. Schumacher, Detection of autologous blood doping with adaptively evaluated biomarkers of doping: a longitudinal blinded study, Transfusion (Paris) 51 (2011) 1707–1715, https://doi.org/10.1111/j.1537-2995.2011.03076.x.
[21] 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) 166–172, https://doi.org/10.1016/j.forsciint.2007.04.001.
[22] D. Roshan, J. Ferguson, C.R. Pedlar, A. Simpkin, W. Wyns, F. Sullivan, J. Newell, A comparison of methods to generate adaptive reference ranges in longitudinal monitoring, PLoS One 16 (2021), e0247338, https://doi.org/10.1371/journal.pone.0247338.
[23] G. Saulière, J. Dedecker, L.-A. Marquet, P. Rochcongar, J.-F. Toussaint, G. Berthelot, Z-scores-based methods and their application to biological monitoring: an example in professional soccer players, Biostatistics 20 (2019) 48–64, https://doi.org/10.1093/biostatistics/kxx044.
[24] C. Saudan, N. Baume, N. Robinson, L. Avois, P. Mangin, M. Saugy, Testosterone and doping control, Br. J. Sports Med. 40 (2006), https://doi.org/10.1136/bjsm.2006.027482 i21–i24.
[25] C. Ayotte, Detecting the administration of endogenous anabolic androgenic steroids, in: D. Thieme, P. Hemmersbach (Eds.), Doping Sports Biochem. Princ. Eff. Anal., Springer, Berlin, Heidelberg, 2010, pp. 77–98, https://doi.org/10.1007/978-3-540-79088-4_4.
[26] WADA, Technical Document – TD2004EAAS, Reporting and evaluation guidance for testosterone, epitestosterone, T/E ratio and other endogenous steroids, 2004. https://www.wada-ama.org/sites/default/files/resources/files/WADA-TD2004-EAAS-Reporting-and-Evaluation-Guidance-for-Testosterone%2C-Epitestosterone%2C-T-E-Ratio-and-Other-Endogenous-Steroids.pdf. (Accessed 21 November 2022). accessed.
[27] T. Piper, H. Geyer, N. Haenelt, F. Huelsemann, W. Schaenzer, M. Thevis, Current insights into the steroidal module of the athlete biological passport, Int. J. Sports Med. 42 (2021) 863–878, https://doi.org/10.1055/a-1481-8683.
[28] E. Eklund, A. Andersson, L. Ekström, A.L. Hirschberg, Urinary steroid profile in elite female athletes in relation to serum androgens and in comparison with untrained controls, Front. Physiol. 12 (2021). https://www.frontiersin.org/articles/10.3389/fphys.2021.702305. (Accessed 21 November 2022).