

# Z-scores-based methods and their application to biological monitoring: an example in professional soccer players

GUILLAUME SAULIÈRE\*

*Institut National du Sport de l'Expertise et de la Performance, Institut de Recherche bioMédicale et d'Épidémiologie du Sport (IRMES), France, Université Paris-Sud Orsay, France and Université Paris Descartes, EA 7329 Paris, France*

guillaume.sauliere@insep.fr

JÉRÔME DEDECKER

*Université Paris Descartes, Sorbonne Paris Cité, Laboratoire MAP5 (UMR 8145), Paris, France*

LAURIE-ANNE MARQUET

*Institut National du Sport de l'Expertise et de la Performance, Institut de Recherche bioMédicale et d'Épidémiologie du Sport (IRMES), France and Université Paris Descartes, EA 7329 Paris, France*

PIERRE ROCHCONGAR†

*Fédération Française de Football, Ligue de Football Professionnel (LFP) Paris, France*

JEAN-FRANCOIS TOUSSAINT

*Institut National du Sport de l'Expertise et de la Performance, Institut de Recherche bioMédicale et d'Épidémiologie du Sport (IRMES), France, Université Paris Descartes, EA 7329 Paris, France and Assistance Publique—Hôpitaux de Paris, Centre d'Investigation en Médecine du Sport, Hôtel-Dieu (CIMS), Paris, France*

GEOFFROY BERTHELOT

*Institut National du Sport de l'Expertise et de la Performance, Institut de Recherche bioMédicale et d'Épidémiologie du Sport (IRMES), France, Université Paris Descartes, EA 7329 Paris, France and REsearch LABoratory for Interdisciplinary Studies (RELAI), Paris, France*

## SUMMARY

The clinical and biological follow-up of individuals, such as the biological passport for athletes, is typically based on the individual and longitudinal monitoring of hematological or urine markers. These follow-ups aim to identify abnormal behavior by comparing the individual's biological samples to an established baseline. These comparisons may be done via different ways, but each of them requires an appropriate

\*To whom correspondence should be addressed.

†Deceased 16 December 2016.

extra population to compute the significance levels, which is a non-trivial issue. Moreover, it is not necessarily relevant to compare the measures of a biomarker of a professional athlete to that of a reference population (even restricted to other athletes), and a reasonable alternative is to detect the abnormal values by considering only the other measurements of the same athlete. Here we propose a simple adaptive statistic based on maxima of Z-scores that does not rely on the use of an extra population. We show that, in the Gaussian framework, it is a practical and relevant method for detecting abnormal values in a series of observations from the same individual. The distribution of this statistic does not depend on the individual parameters under the null hypothesis, and its quantiles can be computed using Monte Carlo simulations. The proposed method is tested on the 3-year follow-up of ferritin, serum iron, erythrocytes, hemoglobin, and hematocrit markers in 2577 elite male soccer players. For instance, if we consider the abnormal values for the hematocrit at a 5% level, we found that 5.57% of the selected cohort had at least one abnormal value (which is not significantly different from the expected false-discovery rate). The approach is a starting point for more elaborate models that would produce a refined individual baseline. The method can be extended to the Gaussian linear model, in order to include additional variables such as the age or exposure to altitude. The method could also be applied to other domains, such as the clinical patient follow-up in monitoring abnormal values of biological markers.

*Keywords:* Biological monitoring; Biological passport; Clinical follow-up; Longitudinal monitoring; Z-score.

## 1. INTRODUCTION

The systematic monitoring of biological variables over time allows for the early detection of physiological or biological changes. Such longitudinal studies are currently used for the identification of atypical biological observations in elite athletes (see in particular [Saugy and others, 2014](#); [Silva and others, 2013](#); [Sottas and others, 2007, 2011](#); [Zorzoli and others, 2014](#); [Zorzoli and Rossi, 2010](#); [Damsgaard and others, 2006](#); [Parisotto and others, 2000](#); [Lobigs and others, 2017b](#)). It is known that the biological markers are impacted by both the frequency and intensity of training, the environment or health issues ([Malcovati and others, 2003](#); [Nikolaidis and others, 2014](#); [Thirup, 2003](#); [Hecksteden and others, 2016b](#); [Lombardi and others, 2011](#); [Sharpe and others, 2002](#)). This biological monitoring allows for determining the baseline value and its associated variability, provided a sufficient follow-up, and a large cohort. Several sport organizations and researchers already developed such approaches in order to establish the biological profile of elite athletes for doping or medical concerns ([Sallet and others, 2008](#); [Malm and others, 2016](#)). A well-known implementation is the athlete biological passport (APB) that is designed to indirectly reveal the effects of doping ([Zorzoli and Rossi, 2010](#); [Robinson and others, 2007](#); [Pottgiesser and others, 2011](#)). It was officially introduced in 2009 and consists of a normalized follow-up of the athlete's hematological parameters. The data generated by the APB can be analyzed using a Bayes procedure ([Sottas and others, 2007](#); [Pottgiesser and others, 2011](#)), which uses both an *a priori* distribution on the individual parameters and the progressive inclusion of individual's values to learn the characteristics of the individual. This Bayes procedure is now investigated in other biomarkers, including the monitoring of muscle recovery ([Hecksteden and others, 2016a](#)). However, [Malcovati and others \(2003\)](#) and [Egger and others \(2016\)](#) underlined that the inter-individual variability is higher than the intra-individual one, suggesting that the use and design of population data for estimating intra-individual variability is an issue. Other concerns were raised by [Banfi and others \(2011\)](#) who pointed out that the APB methodology does not take any environmental seasonality into account. We shall say more on this question, and discuss other possible applications of a Bayes procedure to our context, at the end of Section 5. The aim of this work is to provide a tractable, yet easy to implement, method for detecting abnormal values in a series of measures from the same individual, complementary to biological analyses. We introduce three new methods based on appropriate Z-scores and their multivariate versions to take care of several markers simultaneously

(as suggested by Parisotto *and others*, 2000; Julian *and others*, 2017; Hecksteden *and others*, 2016b; Malm *and others*, 2016). The simplicity of these methods allows for a clear understanding of the inner workings. These methods are evaluated on the 3-year follow-up of five biological markers for elite male soccer players practicing in the French soccer league. In the section below, we describe the methodological context and explain our three methods.

## 2. METHODS

Let  $X_1, \dots, X_n$  be  $n$  independent Gaussian random variables, with  $X_i \sim \mathcal{N}(\mu_i, \sigma^2)$ . We shall first consider a preliminary method (Method 0), to see if the “new” observation  $x_n$  is abnormal, given  $x_1, \dots, x_{n-1}$ . Method 1 generalizes this approach, by testing if one of the observations in the sample is abnormal given the other ones. Method 2 is designed to detect if a series of consecutive observations in the sample are abnormal, given the rest of the sample. For these three methods (Method 0, Method 1, and Method 2), we propose a simple statistic  $T_n$  whose distribution is known under the appropriate null hypothesis, and a decision rule associated with this statistic.

Method 3 is somewhat different, because we want to account for possible seasonality (the behavior of the biomarkers may differ according to the time of year (summer or winter for the data set presented in Section 4.1)). In that case, we split the sample into two subsamples  $X_1, \dots, X_{n_1}$  and  $Y_1, \dots, Y_{n_2}$  with  $n = n_1 + n_2$ , and we denote by  $\mu_{X_i}$  and  $\mu_{Y_j}$  the expectations of  $X_i$  and  $Y_j$ . Here again, we propose a simple statistic  $T_n$  to detect if the observation  $x_n$  is abnormal given the previous observations in the same subsample.

In the following subsections, we describe these methods in detail. For methods 0, 1, and 3, we shall also present a multivariate extension in the case where  $X_1, \dots, X_n$  are  $n$  independent Gaussian random vectors. These extensions allow us to deal with the situation where several biomarkers are systematically measured on the same individual, in order to avoid multi-test issues.

### 2.1. Method 0

To detect if the observation  $x_n$  is abnormal given the mean of the previous observations, the usual way is to consider the statistic

$$T_n = \frac{X_n - \bar{X}_{n-1}}{\hat{\sigma}_{n-1} \sqrt{1 + \frac{1}{n-1}}},$$

where  $\bar{X}_{n-1}$  and  $\hat{\sigma}_{n-1}^2$  are the empirical mean and variance of the sample  $X_1, \dots, X_{n-1}$ , that is

$$\bar{X}_{n-1} = \frac{1}{n-1} \sum_{k=1}^{n-1} X_k \quad \text{and} \quad \hat{\sigma}_{n-1}^2 = \frac{1}{n-2} \sum_{k=1}^{n-1} (X_k - \bar{X}_{n-1})^2.$$

Sottas *and others* (2007) quoted that, for small  $n$ , one cannot use this statistic to detect an abnormal value of  $X_n$  by comparing the observed value  $|t_n|$  of  $|T_n|$  to the quantile  $1 - \alpha/2$  of the  $\mathcal{N}(0, 1)$  distribution. This remark is correct, because for small  $n$ , the distribution of  $T_n$  is far from the  $\mathcal{N}(0, 1)$  distribution. However, this issue can easily be overcome by taking the quantile of the exact distribution of  $T_n$  under the null hypothesis  $H_0 : \mu_1 = \mu_2 = \dots = \mu_n = \mu$ . Indeed, under  $H_0$ , it follows from Cochran’s Theorem that  $\hat{\sigma}_{n-1}$  is independent of the couple  $(X_n, \bar{X}_{n-1})$ . From this simple fact, we easily deduce that, under  $H_0$ , the statistic  $T_n$  has the Student( $n - 2$ ) distribution.

The decision rule is then the following: at level  $\alpha$ , we consider that the value  $x_n$  is abnormal if  $|t_n| > c_{\alpha/2, n-2}$ , where  $c_{\alpha/2, n-2}$  is the quantile of order  $1 - \alpha/2$  of the Student( $n - 2$ ) distribution. This method can be applied as soon as  $n \geq 3$ .

*Remark.* If there are  $N$  different samples (corresponding to  $N$  different individuals) Sharpe *et al.* propose to use the statistic

$$Z_n = \frac{X_n - \bar{X}_{n-1}}{\hat{\sigma}_{\text{uni}} \sqrt{1 + \frac{1}{n-1}}},$$

for each sample  $X_1, \dots, X_n$  (the size  $n$  may not be the same for all samples). Here, the standard error  $\sigma_{\text{uni}}$  is assumed to be the same for all the variables of all the samples, and  $\hat{\sigma}_{\text{uni}}$  is the estimator of  $\sigma_{\text{uni}}$  built from all the variables. In that case, if  $N$  is large, the distribution of  $Z_n$  is close to the  $\mathcal{N}(0, 1)$  distribution, and, at level  $\alpha$ , the quantile  $1 - \alpha/2$  of the  $\mathcal{N}(0, 1)$  can be used to detect abnormal values. However, as quoted in [Sottas and others \(2007\)](#), the assumption that  $\sigma_{\text{uni}}$  is the same for all the variables is unrealistic for many biomarkers (in general, the variance depends on the individuals).

*Multivariate extension.* In this paragraph, we briefly present the multivariate extension of Method 0. Here  $X_1, \dots, X_n$  are  $n$  independent Gaussian  $\mathbb{R}^d$ -valued random vectors, where  $X_i \sim \mathcal{N}(\mu_i, C)$  and  $C$  is assumed to be invertible. Again, we want to detect if the observation  $x_n$  is abnormal given the previous observations. This corresponds to the case where the same  $d$  markers are observed at each medical visit. One could of course apply the previous procedure to each marker separately, but a single decision rule is preferable in order to avoid multi-test issues. In that case, one can use the statistic

$$T_n = \frac{(n-1)}{nd} (X_n - \bar{X}_{n-1})' C_{n-1}^{-1} (X_n - \bar{X}_{n-1}),$$

where

$$\bar{X}_{n-1} = \frac{1}{n-1} \sum_{k=1}^{n-1} X_k \quad \text{and} \quad C_{n-1} = \frac{1}{n-1-d} \sum_{k=1}^{n-1} (X_k - \bar{X}_{n-1})(X_k - \bar{X}_{n-1})'.$$

Under  $H_0 : \mu_1 = \mu_2 = \dots = \mu_n = \mu$ , the statistic  $T_n$  has the Fisher( $d, n - 1 - d$ ) distribution (see Proposition 8.14 in [Eaton, 1983](#)).

## 2.2. Method 1

Given the observations  $x_1, \dots, x_n$  from the sample  $X_1, \dots, X_n$  there seems to be no good reason to consider only the case where the last observation could be abnormal given the previous ones. A much more natural question is: how to detect an abnormal observation among all the observations? Even if one focuses only on the last observation, this question is of interest, because the presence of one (or more) abnormal observation in the past could lead to a wrong decision about the new observation  $x_n$ .

The first way is to iterate Method 0 as follows: one checks if  $x_3$  is abnormal given  $x_1, x_2$ , then one checks if  $x_4$  is abnormal given  $x_1, x_2, x_3$ , and so on up to the last step where one checks if  $x_n$  is abnormal given  $x_1, \dots, x_{n-1}$ . This approach is not satisfactory for at least two reasons: First, the procedure is not invariant by time inversion (the forward and backward procedures can lead to two distinct results). Next, there is a multi-test problem, because  $n - 2$  tests are performed. If each test is performed at level  $\alpha$ , then the level of the series of tests will be greater than  $\alpha$ . Moreover the level correction seems difficult, since these tests are not independent, and the Bonferroni correction is known to be too conservative.

Hence, to answer this question, a global test is preferable. We propose the statistic

$$T_n = \max_{i \in \{1, \dots, n\}} \left| \frac{X_i - \bar{X}_{n,-i}}{\hat{\sigma}_{n,-i} \sqrt{1 + \frac{1}{n-1}}} \right|,$$

where

$$\bar{X}_{n,-i} = \frac{1}{n-1} \sum_{k=1, k \neq i}^n X_k \quad \text{and} \quad \hat{\sigma}_{n,-i}^2 = \frac{1}{n-2} \sum_{k=1, k \neq i}^n (X_k - \bar{X}_{n,-i})^2.$$

This method, based on the maximum of  $n$  (non-independent) variables with Student( $n-2$ ) distribution, can be seen as a straightforward extension of Method 0.

Under the null hypothesis  $H_0 : \mu_1 = \mu_2 = \dots = \mu_n = \mu$ , one can easily check that the distribution  $P_n$  of  $T_n$  does not depend on the unknown parameters  $(\mu, \sigma^2)$ , and can therefore be tabulated.

The decision rule is then the following: at level  $\alpha$ , we decide that there is (at least) one abnormal observation if  $t_n > c_{\alpha,n}$ , where  $c_{\alpha,n}$  is such that  $P_n([c_{\alpha,n}, \infty]) = \alpha$ . This method can be applied as soon as  $n \geq 3$ .

*Multivariate extension.* Let  $X_1, \dots, X_n$  be  $n$  independent Gaussian  $\mathbb{R}^d$ -valued random vectors, where  $X_i \sim \mathcal{N}(\mu_i, C)$  and  $C$  is assumed to be invertible. We want to detect if vector  $x_\ell$  is abnormal given the others. We propose the statistic:

$$T_n = \frac{(n-1)}{nd} \max_{i \in \{1, \dots, n\}} (X_i - \bar{X}_{n,-i})' C_{n,-i}^{-1} (X_i - \bar{X}_{n,-i}),$$

where

$$\bar{X}_{n,-i} = \frac{1}{n-1} \sum_{k=1, k \neq i}^n X_k \quad \text{and} \quad C_{n,-i} = \frac{1}{n-1-d} \sum_{k=1, k \neq i}^n (X_k - \bar{X}_{n,-i})(X_k - \bar{X}_{n,-i})'.$$

Under the null hypothesis  $H_0 : \mu_1 = \mu_2 = \dots = \mu_n = \mu$ , one can easily check that the distribution  $P_n$  of  $T_n$  does not depend on  $\mu$  nor  $C$ , and can therefore be tabulated.

The decision rule is then the following: at level  $\alpha$ , we decide that there is (at least) one abnormal observation if  $t_n > c_{\alpha,n}$ , where  $c_{\alpha,n}$  is such that  $P_n([c_{\alpha,n}, \infty]) = \alpha$ . This method can be applied as soon as  $n \geq d+2$ . Note that, if  $d=1$ , the statistic  $T_n$  is exactly the square of the statistic described before.

### 2.3. Method 2

Recall that, in our setting, the index  $i$  of the observation  $x_i$  represents the order of observations (i.e. the observation  $x_i$  is the  $i$ -th measurement of a biomarker on the same individual). In this context, another interesting question is: how to detect a series of consecutive observations that are abnormal given the rest of the series?

More precisely, we want to know if there is a set  $I = \{i, \dots, j\}$  of consecutive integers, for which the expectation of the  $X_k$ ,  $k \in I$  is different from the expectation of the other variables, that is:  $\mu_k = \mu_\ell = \mu_A$  if  $k, \ell$  do not belong to  $I$  and  $\mu_k = \mu_\ell = \mu_B$  if  $k, \ell$  belong to  $I$ .

To answer this question, we propose the statistic

$$T_n = \max_{I \in \mathcal{I}} \left| \frac{\bar{X}_I - \bar{X}_{\bar{I}}}{\hat{\sigma}_{n,I} \sqrt{\frac{1}{|I|} + \frac{1}{n-|I|}}} \right|,$$

where  $\mathcal{I}$  is the collection of all possible intervals  $I$  included in  $\{1, \dots, n\}$  with length  $1 \leq |I| < n$ ,

$$\bar{X}_I = \frac{1}{|I|} \sum_{k \in I} X_k, \quad \bar{X}_{\bar{I}} = \frac{1}{n-|I|} \sum_{k \notin I} X_k,$$

and

$$\hat{\sigma}_{n,I}^2 = \frac{1}{n-2} \left( \sum_{k \in I} (X_k - \bar{X}_I)^2 + \sum_{k \notin I} (X_k - \bar{X}_{\bar{I}})^2 \right).$$

Under the null hypothesis  $H_0 : \mu_1 = \mu_2 = \dots = \mu_n = \mu$ , one can easily check that the distribution  $P_n$  of  $T_n$  does not depend on the unknown parameters  $(\mu, \sigma^2)$ , and can therefore be tabulated.

The decision rule is then the following: at level  $\alpha$ , we decide that there is (at least) one consecutive series of abnormal observations if  $t_n > c_{\alpha,n}$ , where  $c_{\alpha,n}$  is such that  $P_n([c_{\alpha,n}, \infty[) = \alpha$ . This method can be applied as soon as  $n \geq 3$ .

### 2.4. Method 3

Let us recall the context: we have two series of observations  $x_1, \dots, x_{n_1}$  and  $y_1, \dots, y_{n_2}$  and we want to know if one of these observations is abnormal given the other ones in the same subsample.

We propose the statistic

$$T_n = \max \left\{ \max_{i \in \{1, \dots, n_1\}} \left| \frac{X_i - \bar{X}_{n_1, -i}}{\hat{\sigma}_{X, -i, Y} \sqrt{1 + \frac{1}{n_1 - 1}}} \right|, \max_{j \in \{1, \dots, n_2\}} \left| \frac{Y_j - \bar{Y}_{n_2, -j}}{\hat{\sigma}_{Y, -j, X} \sqrt{1 + \frac{1}{n_2 - 1}}} \right| \right\},$$

where

$$\bar{X}_{n_1, -i} = \frac{1}{n_1 - 1} \sum_{k=1, k \neq i}^{n_1} X_k, \quad \bar{Y}_{n_2, -j} = \frac{1}{n_2 - 1} \sum_{k=1, k \neq j}^{n_2} Y_k,$$

$$\hat{\sigma}_{X, -i, Y}^2 = \frac{1}{n-3} \left( \sum_{k=1, k \neq i}^{n_1} (X_k - \bar{X}_{n_1, -i})^2 + \sum_{k=1}^{n_2} (Y_k - \bar{Y}_{n_2})^2 \right),$$

and

$$\hat{\sigma}_{Y, -j, X}^2 = \frac{1}{n-3} \left( \sum_{k=1}^{n_1} (X_k - \bar{X}_{n_1})^2 + \sum_{k=1, k \neq j}^{n_2} (Y_k - \bar{Y}_{n_2, -j})^2 \right).$$

Under the null hypothesis

$$H_0 : \mu_{X_1} = \mu_{X_2} = \dots = \mu_{X_{n_1}} = \mu_X \text{ and } \mu_{Y_1} = \mu_{Y_2} = \dots = \mu_{Y_{n_2}} = \mu_Y,$$

one can easily check that the distribution  $P_n$  of  $T_n$  does not depend on the unknown parameters  $(\mu_X, \mu_Y, \sigma^2)$ , and can therefore be tabulated.

The decision rule is then the following: at level  $\alpha$ , we decide that there is (at least) one abnormal observation among the two samples if  $t_n > c_{\alpha,n}$ , where  $c_{\alpha,n}$  is such that  $P_n([c_{\alpha,n}, \infty]) = \alpha$ . This method can be applied as soon as  $n_1 \geq 2$  and  $n_2 \geq 2$ .

*Multivariate extension.* Here  $X_1, \dots, X_{n_1}$  and  $Y_2, \dots, Y_{n_2}$  are two independent series of independent Gaussian  $\mathbb{R}^d$ -valued random vectors, with  $X_i \sim \mathcal{N}(\mu_{X_i}, C)$ ,  $Y_j \sim \mathcal{N}(\mu_{Y_j}, C)$  and  $C$  is assumed to be invertible. We want to detect if the vector  $x_k$  or  $y_\ell$  is abnormal given the other ones in the same subsample. We propose the statistic

$$T_n = \max \left\{ \frac{(n_1 - 1)}{n_1 d} \max_{i \in \{1, \dots, n_1\}} (X_i - \bar{X}_{n_1, -i})' C_{X, -i, Y}^{-1} (X_i - \bar{X}_{n_1, -i}), \right. \\ \left. \frac{(n_2 - 1)}{n_2 d} \max_{j \in \{1, \dots, n_2\}} (Y_j - \bar{Y}_{n_2, -j})' C_{Y, -j, X}^{-1} (Y_j - \bar{Y}_{n_2, -j}) \right\},$$

where

$$\bar{X}_{n_1, -i} = \frac{1}{n_1 - 1} \sum_{k=1, k \neq i}^{n_1} X_k, \quad \bar{Y}_{n_2, -j} = \frac{1}{n_2 - 1} \sum_{k=1, k \neq j}^{n_2} Y_k, \\ C_{X, -i, Y} = \frac{1}{n - d - 2} \left( \sum_{k=1, k \neq i}^{n_1} (X_k - \bar{X}_{n_1, -i})(X_k - \bar{X}_{n_1, -i})' + \sum_{k=1}^{n_2} (Y_k - \bar{Y}_{n_2})(Y_k - \bar{Y}_{n_2})' \right),$$

and

$$C_{Y, -j, X} = \frac{1}{n - d - 2} \left( \sum_{k=1}^{n_1} (X_k - \bar{X}_{n_1})(X_k - \bar{X}_{n_1})' + \sum_{k=1, k \neq j}^{n_2} (Y_k - \bar{Y}_{n_2, -j})(Y_k - \bar{Y}_{n_2, -j})' \right).$$

Under the null hypothesis

$$H_0 : \mu_{X_1} = \mu_{X_2} = \dots = \mu_{X_{n_1}} = \mu_X \text{ and } \mu_{Y_1} = \mu_{Y_2} = \dots = \mu_{Y_{n_2}} = \mu_Y,$$

one can easily check that the distribution  $P_n$  of  $T_n$  does not depend on the unknown parameters  $(\mu_X, \mu_Y, C)$  and can therefore be tabulated.

The decision rule is then the following: at level  $\alpha$ , we decide that there is (at least) one abnormal observation among the two samples if  $t_n > c_{\alpha,n}$ , where  $c_{\alpha,n}$  is such that  $P_n([c_{\alpha,n}, \infty]) = \alpha$ . This method can be applied as soon as  $n_1 \geq 2$ ,  $n_2 \geq 2$ , and  $n \geq d + 3$ .

### 3. QUANTILES, POWER, AND ASYMPTOTIC DISTRIBUTION

In the following sections, we use MATLAB R2011b to compute the quantiles and assess the power of the methods. The tests and graphs were done using R v3.3.2.

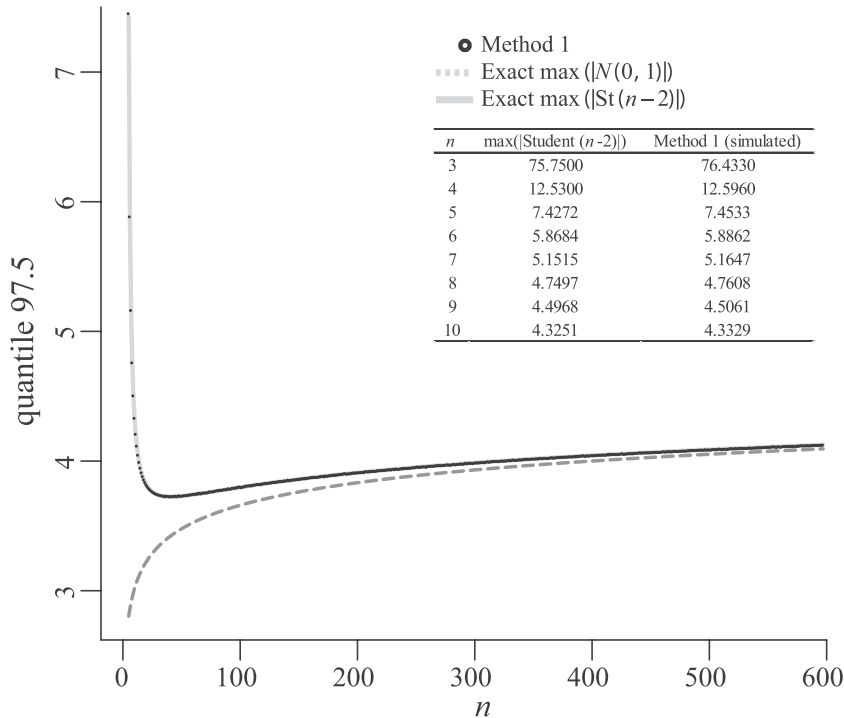


Fig. 1. The levels estimated for Method 1 ( $\alpha = 2.5\%$ ), along with the exact levels for  $\max(|\mathcal{N}(0, 1)|)$  and  $\max(|\text{Student}(n-2)|)$  for  $n = 3$  to 597 observations ( $10^6$  individuals). The levels are detailed in the inset table for the first 10 observations.

### 3.1. Computation of the quantiles

In this section, we explain how to compute the quantiles of the distributions of the statistics  $T_n$  under the null hypothesis, for methods 0, 1, 2, and 3. For Method 0, the quantile of order  $\alpha$  of distribution of the statistic  $|T_n|$  is simply the quantile of order  $1 - \alpha/2$  of the Student( $n - 2$ ) distribution. The statistic  $T_n$  of Method 1 is the maximum of the absolute value of non-independent variables with Student( $n - 2$ ) distribution (respectively Fisher ( $d, n - 1 - d$ ) distribution in the multivariate case). To compute the quantiles of the distribution of this statistic, we use a standard Monte-Carlo procedure with  $20 \times 10^6$  independent experiments, for each level  $\alpha$ . For each  $n$  from 3 to 20, the quantiles are computed from 80% to 99.99% (see [Supplementary material](#) available at *Biostatistics* online). The same method is used for computing the quantiles of the distributions of the statistics of Methods 2 and 3 (please refer to [Supplementary material](#) available at *Biostatistics* online). For Method 1, we also compare the quantiles obtained by simulation to the quantiles of the distribution of the maximum of the absolute values of  $n$  independent variables with Student( $n - 2$ ) distribution. For the quantile 97.5%, we see that the simulated quantiles and the exact quantiles of this approximating distribution are almost identical as soon as  $n > 5$ . Hence, at this level, the exact quantile of the approximating distribution can be used as soon as  $n > 5$ . In Figure 1, we plot the quantile 97.5% obtained by simulation, the exact quantile of the approximating distribution, and also the exact quantiles of the distribution of the absolute values of  $n$  independent random variables with distribution  $\mathcal{N}(0, 1)$ . We see that the convergence to the quantiles based on the Gaussian distribution is quite slow.



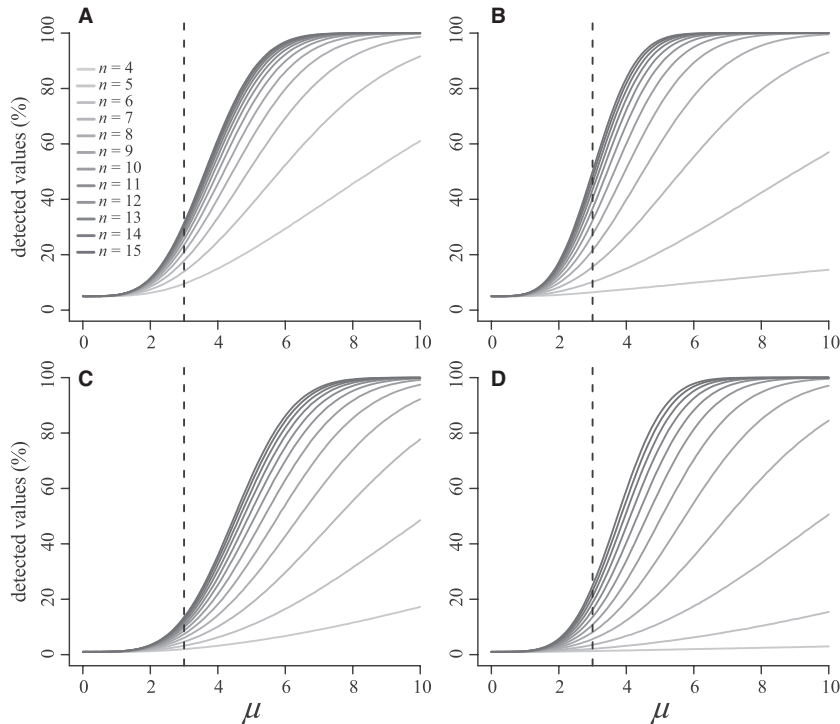


Fig. 2. Frequency of abnormal values detected by Method 1 (A, C) and its multivariate extension (B, D) as a function of  $\mu$  and  $\mu_d$ , for the levels  $\alpha = 5\%$  (A, B) and  $\alpha = 1\%$  (C, D). The dashed line at  $\mu = 3$  is a visual indicator.

### 3.2. Power of the Method 1

In this subsection, we show how Method 1 will detect an abnormal value when all the others are identically distributed. We proceed as follows: we simulate  $1 \times 10^6$  sequences consisting of  $n$  independent Gaussian random variables. The third random variable has distribution  $\mathcal{N}(\mu, 1)$  with  $\mu \geq 0$ , whereas all the other random variables have distribution  $\mathcal{N}(0, 1)$ . The simulations are performed for a number of observations  $n$  ranging from 4 to 15 and  $\mu$  ranging from 0 to 10 with a step of 0.1. We use the same procedure for the multivariate analysis, with  $d = 2$ . The third random vector has distribution  $\mathcal{N}(\tilde{\mu}, C)$ , whereas all the other random vectors have distribution  $\mathcal{N}(0, Id)$ . Here  $\tilde{\mu} = (\mu, \mu)$  with the same possible values of  $\mu$  as in the univariate case, and the covariance matrix  $C$  is given by

$$C = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

The results are given in Figure 2 below.

As expected, the power curve is increasing with  $n$  and  $\mu$ . For the univariate version of Method 1, percentages of abnormal sequences detected for  $n = 9$  and  $\mu$ , respectively, equals to 2, 4, and 6 are 10.53%, 51.06%, and 90.57%. Similar graphs can be obtained for Methods 2 and 3.

### 3.3. Asymptotic result for Methods 1 and 3

In this subsection, we give the asymptotic distribution (as  $n$  goes to infinity) of the (centered and normalized) statistic  $T_n$  of Method 1 under  $H_0$ . At the end, we shall also deal with the statistic  $T_n$  of Method 3. This asymptotic distribution is the Gumbel distribution. As  $n$  goes to infinity, the distribution of  $T_n$  is close to the distribution of the maximum of the absolute value of  $n$  iid random variables with  $\mathcal{N}(0, 1)$  distribution (see Figure 1). Recall the expression of the terms of standardization introduced by Hall (1979):

$$\beta_n = \sqrt{2 \log n} - \frac{\log(4\pi \log n)}{2\sqrt{2 \log n}}.$$

One can prove the following result: under the null hypothesis  $H_0$ , the distribution of the random variable  $\beta_{2n}(T_n - \beta_{2n})$  converges to a Gumbel distribution, that is a distribution whose cumulative distribution function is given by  $\Lambda(x) = \exp(-\exp(-x))$ .

The proof of this proposition is not difficult. The main point is to replace the empirical means and variances in the definition of  $T_n$  by the true values of the parameters  $\mu$  and  $\sigma^2$ . Then, apply a known result for the distribution of the maximum absolute value of  $n$  iid random variables with  $\mathcal{N}(0, 1)$  distribution (see for instance Gasull and others, 2015a). As for the Gaussian case, the convergence of the maximum to the Gumbel distribution is very slow (rate of order  $\log n$ ). So the quantiles of the Gumbel distribution can be used for very large samples only.

Let us consider now the statistic  $T_n$  proposed at the end of Section 2.2 to deal with the multivariate case. Arguing as in the 1D case, we easily see that the asymptotic distribution of  $T_n$  under  $H_0$  is the same as that of the maximum of  $n$  iid random variables  $\xi_1, \dots, \xi_n$  such that  $d\xi_i \sim \chi^2(d)$ . The standard norming sequences are given for instance in the book by Embrechts and others (2013) (see p. 155). It follows that, under  $H_0$  and for  $d \geq 2$ ,  $d(T_n - b_n)/2$  converges in distribution as  $n \rightarrow \infty$  to the Gumbel distribution, where

$$b_n = \frac{2}{d} \left( \log n + \frac{(d-2)}{2} \log \log n - \log \Gamma \left( \frac{d}{2} \right) \right).$$

We refer to the article by Gasull and others (2015b), Section 5, for more accurate norming sequences.

The same proofs can be done for the statistic  $T_n$  of Method 3, with straightforward modifications. In the 1D case, we find that, under  $H_0$ , the distribution of the random variable  $\beta_{2n}(T_n - \beta_{2n})$  converges to a Gumbel distribution, provided that  $\min(n_1, n_2) > C(\log(n))^a$  for some  $C > 0$  and some  $a > 1$ . In the multivariate case (see the end of Section 2.4), the distribution of  $d(T_n - b_n)/2$  converges under  $H_0$  to a Gumbel distribution under the same condition on  $\min(n_1, n_2)$ . Note that this last restriction on  $n_1, n_2$  is very mild and should be realized in all practical situations.

## 4. APPLICATION TO SOCCER PLAYERS DATA

### 4.1. Description of the data base

After obtaining the approval of the Institutional Ethics Committee, our three methods were applied to a database of elite soccer players. It consists of five typical biological markers from 2577 male soccer players from the French elite leagues 1 and 2. Markers include concentrations of ferritin ( $\mu\text{mol/L}$ ), serum iron ( $\mu\text{mol/L}$ ), hemoglobin (g/L), erythrocytes (T/L), and hematocrit levels (%). The biomarkers were collected every 6 months in July/August and in January/March from 2006 to 2012 for a total of 12 collections. The large interval between two measures (around 6 months) allowed for independent sampling (Sharpe and others, 2006). We kept the players that had at least 5 measurements over the 12 possible

Table 1. Percentage of individual series significantly non-normal with the Shapiro test and corresponding mean of  $P$ -values. And results of the sign tests. (L) indicates a log-transformation of the data

Marker	% of non-normal series	Mean $P$ -value of Shapiro tests	$P$ -value of the sign test
Ferritin (L)	7.05	0.47	0.804
Serum iron (L)	7.67	0.49	0.703
Hemoglobin	6.91	0.46	$3 \times 10^{-8}$
Erythrocytes	6.53	0.48	$3 \times 10^{-8}$
Ferritin	5.25	0.48	$1 \times 10^{-10}$

measurements, totalizing: 757 players for ferritin and serum iron, 799 (erythrocytes and hematocrit), and 807 for hemoglobin because of the high number of transfers between clubs, injuries, and the progressive inclusion of new clubs in the elite league. Moreover, a technical issue with the sampling instrument resulted in the loss of the data in the markers of the 2009 July/August collection. According to [Custer and others \(1995\)](#) and [Tufts and others \(1985\)](#) the measure of the ferritin and serum iron have a log-normal distribution, so we take the logarithm of the observations for these two biomarkers.

#### 4.2. Preliminary analysis

The individual series of biomarkers must comply with the conditions described in Section 2. Regarding the assumption of normality: for each individual and each biomarker, we use the Shapiro test to confirm that it is not unrealistic to assume that the observations are drawn from a normal sample.

We note that, for each biomarker, the percentage of non-normal sequences detected by the test is not significantly different from the 5% level (that is the false-discovery rate of the test under the null hypothesis) (see Table 1). Regarding the assumption of equidistribution for Methods 1 and 2, it is known that the training intensity and frequency, the environment, and the period are among the factors inducing a possible variability in biological markers ([Malcovati and others, 2003](#); [Thirup, 2003](#)). The impact of seasonal changes on hematologic markers has already been investigated in both the general and athletic population. However, some uncertainty remains, especially regarding hematocrit as reported in [Malcovati and others \(2003\)](#), [Meyer and Meister \(2011\)](#), and [Thirup \(2003\)](#), opposing [Heisterberg and others \(2013\)](#). Hence, to see if the period of the year has a significant influence on the biomarkers, we chose to perform a sign test based on the successive differences  $D_i = X_{S,i} - X_{W,i}$ , where  $X_{S,i}$  corresponds to the measure in summer of year  $i$ , and  $X_{W,i}$  corresponds to the measure in winter of year  $i$ . We collect all possible  $D_i$ , for all individuals and perform the sign test on this large data set. This is possible because the sign test is very robust, and not influenced by the inter-variability between individuals. All that matters for its application is that the variables  $1_{D_i > 0}$  are *iid*, which is certainly true once we admit that a difference of 6 months is enough for the independency. Under the null hypothesis, the period of the year has no influence on the biomarker, the distribution of the number of positive  $D_i$  follows a binomial distribution  $B(N, 0.5)$  ( $N$  being the total number of  $D_i$  for all the individuals).

The sign tests give very small  $P$ -values for hemoglobin, erythrocytes and hematocrit, meaning that there is a significant difference between summer and winter for these three biomarkers (see Table 1). Therefore, only the ferritin and serum iron are eligible for Methods 1 and 2. For the other markers, the third method can be applied.

For the two markers ferritin and serum iron, we see that the empirical distribution of  $T_n$  (Method 1) is close to the theoretical one under  $H_0$  (see Figure of [Supplementary material](#) available at *Biostatistics*

Table 2. Number of abnormal series for each biomarker and method ( $\alpha = 5\%$ )

	Ferritin	Serum iron	Red blood cells	Hemoglobin	Hematocrit	Total
Method 1						
Usual ( $n$ )	715	695	—	—	—	1410
Unusual ( $n$ )	42	62	—	—	—	104
Unusual (%)	5.55	8.19	—	—	—	6.87
(Multivariate)						
Usual ( $n$ )		661	—	—	—	661
Unusual ( $n$ )		49	—	—	—	49
Unusual (%)		6.90	—	—	—	6.9
Method 2						
Usual ( $n$ )	674	707	—	—	—	1381
Unusual ( $n$ )	83	50	—	—	—	133
Unusual (%)	10.96	6.61	—	—	—	8.78
Method 3						
Usual ( $n$ )	—	—	577	588	593	1758
Unusual ( $n$ )	—	—	41	38	35	114
Unusual (%)	—	—	6.63	6.07	5.57	6.09

online). This confirms that the quantiles of the theoretical distribution of  $T_n$  can be used to detect abnormal values on our dataset.

### 4.3. Results

The three methods are used on the longitudinal follow-up of five biological markers collected from elite soccer players. Some series are log-transformed to ensure that the variables were normally distributed (see Table 1), and we have carefully taken care of the seasonality (detected by the sign test with a very small  $P$ -value, see Table 1 and Section 4.2). As stated previously, it is not clear whether the hematocrit marker is impacted by the seasonal change but our results show a seasonal variability for hemoglobin, erythrocytes, and hematocrit (see Table 1). Hence, we use Method 3 for these three markers, although these variations are possibly linked to the geographical trip of individuals due to training or competition reasons. According to these preliminary analyses, we apply Method 1 (and its multivariate extension) and Method 2 to the ferritin and the serum iron, and Method 3 to erythrocytes, hemoglobin, and hematocrit. We keep the individuals with at least 3 values per season for the Method 3. The frequency of abnormal series detected by the procedures are reported in Table 2 for a level  $\alpha = 5\%$ .

Most of the results are not too far from the expected false-discovery rate: between 5% and 6.7% for Ferritin/Method 1, Serum iron/Method 2, and all three others biomarkers for Method 3; 8.19% for Serum Iron/Method 1. However, the percentage of abnormal subseries for Ferritin/Method 2 is close to 11%, hence significantly different from the expected level  $\alpha = 5\%$ . We cannot directly assert that these differences are related to a pathological behavior; for instance it could be explained by a slight departure from the Gaussian distribution (7.05% of the Ferritin series were detected to be non-normal by the Shapiro test at level 5%). Thus, further analyses are required to shed light on these series. We find the same percentage of abnormal values for the multivariate version of Method 1 as for the two univariate procedures (6.9% in both cases). In Figure 3 below, we give some examples of abnormal observations detected by the three methods, for different biomarkers.

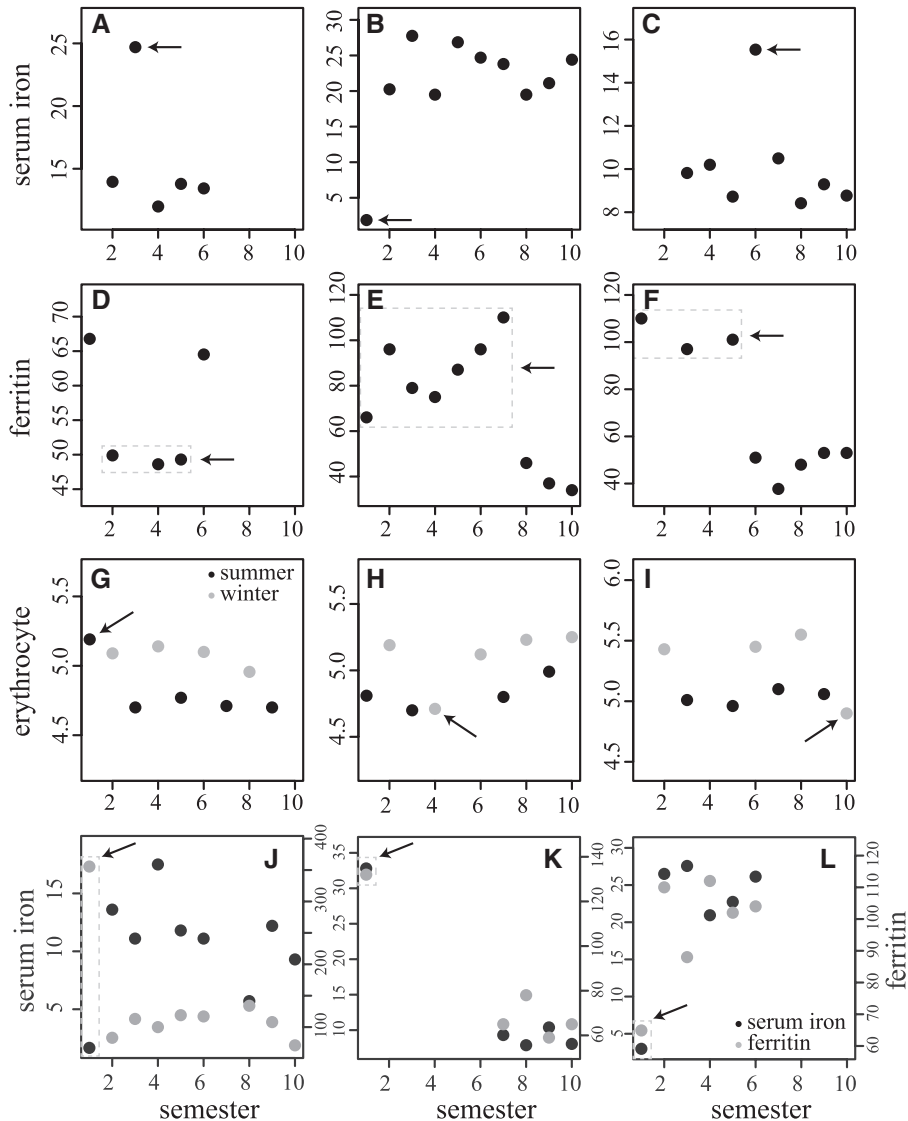


Fig. 3. Examples of results for the Method 1 (A–C), Method 2 (D–F), and Method 3 (G–I) in different individuals for the serum iron, ferritin and erythrocyte values. Panels (J–L) show some results for the multivariate version of Method 1.

## 5. DISCUSSION

This article introduces three Z-scores-based methods for detecting abnormal values in a series of biological measures from an individual's sample. They allow for assessing the individual baseline while taking into account the seasonal changes that alter the values of biomarkers. The multivariate approach is also developed in order to avoid multi-test issues and to take care of the possible correlations between biomarkers. Finally, we make extensive simulations to obtain the power-curve of our procedures for certain points of the alternative hypothesis. This power-curve could be used to compare our procedures to other methods.

Unfortunately, such power curves are not available in the article by [Sottas and others \(2007\)](#), and it is somewhat unclear how they could be produced. Our methods are tested on the follow-up of elite athletes and we found the results in accordance with the expected false-discovery rate, in most of the cases. Our methods can be extended to the Gaussian linear model, but the quantile tables have to be computed for each new design. A possible alternative would be to compute an approximate  $p$ -value by using Monte Carlo simulations.

The longitudinal values of athlete's hematological markers were reported to differ from those of the general population. Several previous works have shown that biological values are affected by the player's position, level, ethnic origin, and age among other parameters. As a result, [Malcovati and others \(2003\)](#) suggested focusing on intra-variability only. [Sharpe and others \(2006\)](#) and [Sottas and others \(2007\)](#) introduced two interesting approaches to identify outliers in longitudinal follow-ups. However, [Sharpe and others \(2006\)](#) assume that the variance of the random variables is the same for all individuals, which is not relevant in many cases. The method proposed by [Sottas and others \(2007\)](#) is convincing and provides good results, but the need of an extra population makes it rather complicated to use (a large relevant population has to be used for each different biomarker). In contrast, our methods are particularly simple, and can be tabulated, and then applied to all biomarkers, provided the distribution is Gaussian (up to a known deterministic transformation).

[Bejder and others \(2016\)](#) showed that acute hyperhydration reduces the sensitivity of the Bayes APB procedure for usual blood markers, and this could affect the sensitivity of our methods as well. However, [Lobigs and others \(2017a\)](#) proposed new plasma volume biomarkers that should help to correct this effect.

To conclude this section, let us discuss the relevance of a hierarchical Bayes procedure in our context. The following are elements of discussion:

- First, for a Bayes procedure, one needs to choose an *a priori* distribution on the parameters  $(\mu, \sigma^2)$ . This can be done by using an extra population, as in [Sottas and others \(2007\)](#), who fit a parametric model to the parameters. If this extra population is not available, but there are many individuals in the study (as in our soccer players database), one can select a subpopulation to perform this first estimation step. Note however, the methods described in the present article can be applied to one single individual without the help of an extra population.
- Once we know the *a priori* distribution, we have access via the Bayes rule to the global distribution of the variables (in the Bayesian framework, the distribution  $N(\mu, \sigma^2)$  is the conditional distribution of the variables  $X_i$  given  $(\mu, \sigma^2)$ ). Hence, we can test if the observations are abnormal with respect to this distribution. However, it is important to point out that this is not what we have in mind: we want to test if, for a given individual, some observations are abnormal; this can be done efficiently only if the characteristics of the individuals are taken into account. If we use the global distribution of the variables, we will primarily detect individuals that are different from the rest of the population (because the observations of the bio-markers of these individuals will very likely be abnormal with respect to the global distribution). But we are not interested to know if an individual is different from the rest of the population (it is irrelevant for professional athletes).
- [Sottas and others \(2007\)](#) used an intermediate approach. Thanks to the Bayes rule again, they compute the conditional distribution of  $X_n$  given  $X_1, \dots, X_{n-1}$ , and they use the quantiles of this distribution to see if the observation  $x_n$  is abnormal or not. This procedure is satisfactory (even if it requires an extra population for each biomarker to determine an appropriate *a priori* distribution). It is similar to our Method 0, but it does not answer the question: is there at least one abnormal value in the whole sequence of measurements of a single individual? One way to answer is to perform a sequential procedure, testing if  $x_2$  is abnormal given  $x_1$ , if  $x_3$  is abnormal given  $x_1, x_2$ ,

and so on . . . but this leads to a non-trivial multi-test problem, as already explained at the beginning of Section 2.2.

To summarize, our methods allow for a single-detection rule, thus avoiding multi-test issues that would occur by testing if each new value is abnormal. In the univariate case, Method 1 can be applied with at least three observations per individual. This is of course a (rather mild) restriction of the method but, on another hand, it can be applied to any sequence of independent Gaussian outcomes, without the help of an extra cohort to determine an appropriate *a priori* distribution.

## 6. CONCLUSION

We propose three methods to detect abnormal values or subseries in longitudinal follow-ups. These methods are simple and easy to implement, and their significance levels can be easily computed via Monte Carlo simulations. They can be applied to independent and normally distributed random variables, for series with at least three observations. This last point does not seem to be a strong limitation, since today's follow-ups often include more than three observations, and the sampling rate is increasing over the years. We apply these methods to a large dataset, and we detect some abnormal values (some of them being false positives). Additional investigations should be carried out by the medical staff in order to shed light on abnormal follow-ups. The proposed methods could be improved to include additional periodic environmental effects, such as training at altitude. Practical applications include detecting abnormal values in elite athletes' follow-ups for clinical or anti-doping purposes. It would also be useful in detecting pathological values in clinical follow-ups of patients based on individual rather than population-based thresholds.

## SUPPLEMENTARY MATERIAL

Supplementary material is available at <http://biostatistics.oxfordjournals.org>.

## ACKNOWLEDGMENTS

This article is dedicated to the work and commitment of Pierre Rochcongar (1947–2016). We deeply thank the Association des Médecins de Club de Football Professionnel represented by Hervé Caumont (MD) and Emmanuel Orhant (MD). We also thank Lydie Jorda and Jean Lagrange for the preliminary analyses and Stacey Johnson for reviewing the manuscript. *Conflict of Interest*: None declared.

## FUNDING

The biomarker collections and analyses were supported by the Ligue de Football Professionnel.

## REFERENCES

- BANFI, G., LOMBARDI, G., COLOMBINI, A. AND LIPPI, G. (2011). Analytical variability in sport hematology: its importance in an antidoping setting. *Clinical Chemistry and Laboratory Medicine* **49**, 779–782.
- BEJDER, J., HOFFMANN, M. F., ASHENDEN, M., NORDSBORG, N. B., KARSTOFT, K. AND MØRKEBERG, J. (2016). Acute hyperhydration reduces athlete biological passport off-hr score. *Scandinavian Journal of Medicine & Science in Sports* **26**, 338–347.
- CUSTER, E. M., FINCH, C. A., SOBEL, R. E. AND ZETTNER, A. (1995). Population norms for serum ferritin. *The Journal of Laboratory and Clinical Medicine* **126**, 88–94.

- DAMSGAARD, R., MUNCH, T., MORKEBERG, J., MORTENSEN, S. P. AND GONZALEZ-ALONSO, J. (2006). Effects of blood withdrawal and reinfusion on biomarkers of erythropoiesis in humans: implications for anti-doping strategies. *Haematologica* **91**, 1006–1008.
- EATON, M. L. (1983). *Multivariate Statistics: A Vector Space Approach*. New York, NY: John Wiley & Sons, Inc., p.512.
- EGGER, F., MEYER, T. AND HECKSTEDEN, A. (2016). Interindividual variation in the relationship of different intensity markers—a challenge for targeted training prescriptions. *PLoS One* **11**, e0165010.
- EMBRECHTS, P., KLÜPPELBERG, C. AND MIKOSCH, T. (2013). *Modelling Extremal Events: For Insurance and Finance*, Volume 33. Berlin Heidelberg New-York: Springer Science & Business Media.
- GASULL, A., JOLIS, M. AND UTZET, F. (2015a). On the norming constants for normal maxima. *Journal of Mathematical Analysis and Applications* **422**, 376–396.
- GASULL, A., LÓPEZ-SALCEDO, J. A. AND UTZET, F. (2015b). Maxima of gamma random variables and other weibull-like distributions and the lambert W function. *Test* **24**, 714–733.
- HALL, P. (1979). On the rate of convergence of normal extremes. *Journal of Applied Probability* **16**, 433–439.
- HECKSTEDEN, A., PITSCH, W., JULIAN, R., PFEIFFER, M., KELLMANN, M., FERRAUTI, A. AND MEYER, T. (2016a). A new method to individualize monitoring of muscle recovery in athletes. *International Journal of Sports Physiology and Performance*, 1–25.
- HECKSTEDEN, A., SKORSKI, S., SCHWINDLING, S., HAMMES, D., PFEIFFER, M., KELLMANN, M., FERRAUTI, A. AND MEYER, T. (2016b). Blood-borne markers of fatigue in competitive athletes—results from simulated training camps. *PloS One* **11**, e0148810.
- HEISTERBERG, M. F., FAHRENKRUG, J., KRUSTRUP, P., STORSKOV, A., KJÆR, M. AND ANDERSEN, J. L. (2013). Extensive monitoring through multiple blood samples in professional soccer players. *The Journal of Strength & Conditioning Research* **27**, 1260–1271.
- JULIAN, R., MEYER, T., FULLAGAR, H. H. K., SKORSKI, S., PFEIFFER, M., KELLMANN, M., FERRAUTI, A. AND HECKSTEDEN, A. (2017). Individual patterns in blood-borne indicators of fatigue—trait or chance. *Journal of Strength and Conditioning Research* **31**, 608–619.
- LOBIGS, L. M., SOTTAS, P.-E., BOURDON, P. C., NIKOLOVSKI, Z., EL-GINGO, M., VARAMENTI, E., PEELING, P., DAWSON, B. AND SCHUMACHER, Y. O. (2017a). A step towards removing plasma volume variance from the athlete’s biological passport: the use of biomarkers to describe vascular volumes from a simple blood test. *Drug Testing and Analysis*.
- LOBIGS, L. M., SOTTAS, P.-E., BOURDON, P. C., NIKOLOVSKI, Z., EL-GINGO, M., VARAMENTI, E., PEELING, P., DAWSON, B. AND SCHUMACHER, Y. O. (2017b). The use of biomarkers to describe plasma-, red cell-, and blood volume from a simple blood test. *American Journal of Hematology* **92**, 62–67.
- LOMBARDI, G., RICCI, C. AND BANFI, G. (2011). Effects of winter swimming on haematological parameters. *Biochimica Medica* **21**, 71–78.
- MALCOVATI, L., PASCUTTO, C. AND CAZZOLA, M. (2003). Hematologic passport for athletes competing in endurance sports: a feasibility study. *Haematologica* **88**, 570–581.
- MALM, C. B., KHOO, N. S., GRANLUND, I., LINDSTEDT, E. AND HULT, A. (2016). Autologous doping with cryopreserved red blood cells—effects on physical performance and detection by multivariate statistics. *PloS One* **11**, e0156157.
- MEYER, T. AND MEISTER, S. (2011). Routine blood parameters in elite soccer players. *International Journal of Sports Medicine* **32**, 875–881.
- NIKOLAIDIS, P., ZIV, G., LIDOR, R. AND ARNON, M. (2014). Inter-individual variability in soccer players of different age groups playing different positions. *Journal of Human Kinetics* **40**, 213–225.
- PARISOTTO, R., GORE, C. J., EMSLIE, K. R., ASHENDEN, M. J., BRUGNARA, C., HOWE, C., MARTIN, D. T., TROUT, G. J. AND HAHN, A. G. (2000). A novel method utilising markers of altered erythropoiesis for the detection of recombinant human erythropoietin abuse in athletes. *Haematologica* **85**, 564–572.



- POTTGIESSER, T., SOTTAS, P.-E., ECHELER, T., ROBINSON, N., UMHAU, M. AND SCHUMACHER, Y. O. (2011). Detection of autologous blood doping with adaptively evaluated biomarkers of doping: a longitudinal blinded study. *Transfusion* **51**, 1707–1715.
- ROBINSON, N., SOTTAS, P.-E., MANGIN, P. AND SAUGY, M. (2007). Bayesian detection of abnormal hematological values to introduce a no-start rule for heterogeneous populations of athletes. *Haematologica* **92**, 1143–1144.
- SALLET, P., BRUNET-GUEDJ, E., MORNEX, R. AND BAVEREL, G. (2008). Study of a new indirect method based on absolute norms of variation to detect autologous blood transfusion. *International Journal of Hematology* **88**, 362.
- SAUGY, M., LUNDBY, C. AND ROBINSON, N. (2014). Monitoring of biological markers indicative of doping: the athlete biological passport. *British Journal of Sports Medicine* **48**, 827–832.
- SHARPE, K., ASHENDEN, M. J. AND SCHUMACHER, Y. O. (2006). A third generation approach to detect erythropoietin abuse in athletes. *Haematologica* **91**, 356–363.
- SHARPE, K., HOPKINS, W., EMSLIE, K. R., HOWE, C., TROUT, G. J., KAZLAUSKAS, R., ASHENDEN, M. J., GORE, C. J., PARISOTTO, R. AND HAHN, A. G. (2002). Development of reference ranges in elite athletes for markers of altered erythropoiesis. *Haematologica* **87**, 1248–1257.
- SILVA, J. R., REBELO, A., MARQUES, F., PEREIRA, L., SEABRA, A., ASCENSÃO, A. AND MAGALHÃES, J. (2013). Biochemical impact of soccer: an analysis of hormonal, muscle damage, and redox markers during the season. *Applied Physiology, Nutrition, and Metabolism* **39**, 432–438.
- SOTTAS, P.-E., BAUME, N., SAUDAN, C., SCHWEIZER, C., KAMBER, M. AND SAUGY, M. (2007). Bayesian detection of abnormal values in longitudinal biomarkers with an application to t/e ratio. *Biostatistics* **8**, 285.
- SOTTAS, P.-E., ROBINSON, N., RABIN, O. AND SAUGY, M. (2011). The athlete biological passport. *Clinical Chemistry* **57**, 969–976.
- THIRUP, P. (2003). Haematocrit. *Sports Medicine* **33**, 231–243.
- TUFTS, D. A., HAAS, J. D., BEARD, J. L. AND SPIELVOGEL, H. (1985). Distribution of hemoglobin and functional consequences of anemia in adult males at high altitude. *The American Journal of Clinical Nutrition* **42**, 1–11.
- ZORZOLI, M., PIPE, A., GARNIER, P. Y., VOUILLAMOZ, M. AND DVORAK, J. (2014). Practical experience with the implementation of an athlete's biological profile in athletics, cycling, football and swimming. *British Journal of Sports Medicine* **48**, 862–866.
- ZORZOLI, M. AND ROSSI, F. (2010). Implementation of the biological passport: the experience of the international cycling union. *Drug Testing and Analysis* **2**, 542–547.

[Received February 8, 2017; revised July 28, 2017; accepted for publication August 21, 2017]