Check for updates

ABPS: An R Package for Calculating the Abnormal Blood Profile Score

Frédéric Schütz [1,2]* and Alix Zollinger [1]

[1] Bioinformatics Core Facility, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland, [2] Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland

The Abnormal Blood Profile Score (ABPS) is used to identify blood doping in sport. It combines seven hematological markers, including hemoglobin level, reticulocytes percent, and haematocrit level, using two different machine learning algorithms in order to create a single score that has a better ability to identify doping than each parameter taken alone. The resulting score allows the detection of several types of doping using a single score and is part of the current Athlete Biological Passport program managed by World Anti-Doping Agency (WADA). We describe ≪ ABPS ≫, an R package that allows the calculation of this score. This is the first software implementation calculating this score that is released publicly. The package also contains functions to calculate the OFF-score (another score used for detection of doping), as well as several test datasets. The package is useful for laboratories conducting anti-doping analyses and for researchers working on anti-doping research projects. In particular, it has been successfully used in projects estimating the prevalence of blood doping.

Keywords: code:R, anti-doping, ABPS, blood doping, machine learning

INTRODUCTION

The Abnormal Blood Profile Score (ABPS) is one of several tools used to identify blood doping in elite athletes. Originally developed at the Swiss Laboratory for Doping Analyses (LAD) in Lausanne (Sottas et al., 2006), it combines 7 hematological markers (reticulocytes percent, hemoglobin level, haematocrit level, red blood cell count, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration) into a single score. According to the original ABPS publication, this combined score is more sensitive (for the same level of specificity) to doping than any of these blood parameters taken separately; in particular, the ABPS alone allows the detection of several types of blood doping (Sottas et al., 2006). ABPS is based on two different classification techniques, a naive Bayesian classifier and an SVM (Support Vector Machine). The two models were trained using a database of 591 blood profiles (including 402 control samples from clean athletes and 189 samples of athletes who abused of an illegal substance); the two scores were then combined using a method called ≪ ensemble averaging ≫ to obtain the final score. The resulting ABPS values are unitless; based on a large collection of test results (>3,000 samples), they typically range between −2.35 and 1 (95% confidence interval, with a mean of −0.67 and a standard deviation of 0.87); more detailed reference distributions and the effect of different factors (sex and age of athlete, type of sport, etc) on the score is described elsewhere (Robinson et al., 2018). While ABPS is not a primary marker of doping, it has been used as corroborative evidence (CAS, 2012) and it is part of the Athlete Biological Passport (ABP) program managed by the World Anti-Doping Agency (WADA; Sottas et al., 2011a). Assuming that the testing was conducted according to the WADA guidelines (WADA, 2018), scores between 0 and 1 indicate a possible suspicion of doping (CAS, 2012); a score above 1 should be found only in 1 in 1,000 clean male athletes.

Several versions of ABPS scoring have been developed over the years, using up to 12 different blood parameters (Sottas et al., 2010); however, the most commonly used version is based on the 7 parameters described above, which can be obtained "on-line" on a portable analyzer. The 7 parameters were selected following a comparison of the performances of models with different numbers of parameters. The original ABPS implementation was created using the Matlab programming language, and WADA maintains a Java version that is used within the ADAMS (Anti-Doping Administration and Management System) database and the Athlete Biological Passport program. However, these versions are only available to anti-doping organizations recognized by WADA.

THE R PACKAGE ABPS

The ABPS package allows users to compute the Abnormal Blood Profile Score using the R statistical software (R Core Team, 2017). It is the first such software implementation that is released publicly. The R software was implemented using the original Matlab software as a reference; in particular, no new model fitting was performed, as the original parameters calculated in Matlab for the naive Bayesian and the SVM classifiers were reused in the R code. As a result, the package does not require the original database of blood profiles, which is not publicly available. The concordance of results obtained using the Matlab software and the R implementation was checked using a series of test data, some of which are included in the package (see section ≪ datasets ≫ below for more details).

The ABPS function available in the package requires the user to provide values for the seven hematological markers for one or several samples, and will then calculate and return the corresponding score or scores. The markers can be specified either as a single data frame (the basic structure for managing data in R) containing the seven parameters, or by specifying separately the seven following variables (the expected units are indicated): HCT [haematocrit level, in [%]], HGB [the hemoglobin level, in [g/dL]], MCH [the mean corpuscular hemoglobin, in [pg]], MCHC [the mean corpuscular hemoglobin concentration, in [g/dL]], MCV [the Mean corpuscular volume, in [fL]], RBC [the red blood cell count, in $[10^6/\mu L]$], RETP [the reticulocytes percent, in [%]]. A short example of use is shown on **Figure 1**; detailed information and examples of use are provided in the help page of the function.

The package also provides a function for calculating the OFF-score (Gore et al., 2013), another indirect indicator of blood doping. Also called "OFF-hr score" or "stimulation index" (Zorzoli, 2011), the OFF-score is part of the Athlete Biological Passport and is routinely used to identify athletes who use a substance prohibited by anti-doping rules (CAS, 2016). It combines the hemoglobin level [HGB, [g/dL]] with the

reticulocytes percent (RETP) using the formula

$$OFF - score = HGB \times 10 - 60 \times \sqrt{RETP}$$

The original publication (Gore et al., 2013) provides thresholds for different populations; OFF-score values typically range between 85 and 95 and in the worst case scenario (a male athlete living at low altitude), values over 133 are considered to be evidence of doping (Marrocco et al., 2012). Note that the original publication assumed that the hemoglobin level is specified in g/L; as hemoglobin levels are generally specified in g/dL (for example, in the ADAMS database) and to ensure coherency with other functions within the package, our OFF-score function also expects hemoglobin levels to be specified in g/dL, as for the ABPS calculation, and it will then convert the units internally (as shown by the presence of a multiplication by 10 in the formula above). The OFF-score function will emit a warning if the units seem wrong.

DATASETS

In addition, the ABPS package provides several datasets linked to blood doping. The first dataset, named ≪ bloodcontrol ≫, contains blood parameter data measured on 13 individuals, provided by the Swiss Laboratory for Doping Analyses (LAD) in Lausanne. These samples are assumed to represent normal population, and were not used for fitting the original ABPS model.

The second dataset, named ≪ blooddoping ≫ contains a series of 13 measurements of blood parameters conducted over several years on a female athlete who was later convicted of repeated doping on the basis of this data (CAS, 2012). The published data contains some errors (in particular, some values were swapped between samples) which were discovered while creating the package and confirmed by WADA. The package provides a corrected version of the dataset, as well as a detailed list of the changes that were made.

For both datasets, the OFF-scores and ABPS values provided were computed by WADA software. The R package can thus be tested by comparing the results it produces with the ones computed by WADA.

DOCUMENTATION AND TESTS

The package comes with extensive documentation, in particular help pages, detailing both the available functions and datasets, as well as a set of unit tests which verify that the code works as it should and that it yields the expected results on the provided datasets. The source code is commented in details and the package passes all the checks provided by R without any error, warning or note.

PACKAGE USAGE

The functions provided by the package can be used by laboratories conducting anti-doping analyses who may want to calculate ABPS values based on measurements of blood

---

Abbreviations: ABPS, Abnormal Blood Profile Score; ADAMS, Anti-Doping Administration and Management System; LAD, Swiss Laboratory for Doping Analyses; WADA, World Anti-Doping Agency.

```
> install.packages("ABPS")        # Install the package from CRAN
> library(ABPS)                    # Load the ABPS library
> OFFscore(HGB=14.6, RETP=0.48)    # Calculate one OFF-score value
[1] 104.4308

> # Calculate one ABPS value
> ABPS(HCT=43.2, HGB=14.6, MCH=31.1, MCHC=33.8, MCV=92.1, RBC=4.69, RETP=0.48)
[1] -0.5398784

> data(bloodcontrol)              # Load the "bloodcontrol" data
> head(bloodcontrol)              # Show the first rows of the dataset
   HCT  HGB   IRF  MCH MCHC  MCV  RBC RDW.SD   RETC RETP OFFscore  ABPS
1 48.3 17.1   4.4 28.8 35.4 81.3 5.94   37.4 0.0606 1.02   110.40 -1.12
2 48.8 16.4   9.6 28.0 33.6 83.3 5.86   38.0 0.0703 1.20    98.30 -0.88
3 45.9 15.8  10.0 29.0 34.4 84.2 5.45   38.6 0.0818 1.50    84.50 -0.62
4 47.4 16.0   4.9 28.0 33.8 82.9 5.72   38.1 0.0686 1.20    94.30 -1.24
5 43.8 15.4   4.2 28.4 35.2 80.7 5.43   36.7 0.0559 1.03    93.11 -2.06
6 45.7 15.5   7.7 28.0 33.9 82.6 5.53   37.9 0.0630 1.14    90.90 -1.64

> ABPS(bloodcontrol)     # Calculate ABPS values for the whole dataset
 [1] -1.1248201 -0.8782518 -0.6215326 -1.2370152 -2.0602383 -1.6376824
 [7] -2.1296246 -0.8916582 -1.7010020 -1.4032758 -1.3123875 -0.7967497
[13]  0.1579954
```

FIGURE 1 | A sample session using the ABPS package within R. During this session, the package is installed, loaded, the ≪ bloodcontrol ≫ dataset is loaded, and some calculations of ABPS and OFF-score are performed.

parameters they have performed, or by people interested in anti-doping who want to examine and understand how the ABPS scheme works, and how sensitive it is to changes in one blood parameter or another.

However, the most likely use of the package (and the primary motivation for its creation) is in anti-doping research. As an example, the ABPS was used in the past to estimate the prevalence of doping in different populations of athletes (Sottas et al., 2011b). In this context, ABPS allows the researcher to work with a single, combined, parameter instead of seven different markers. The comparison of its distribution for different groups of interest can be used to estimate the prevalence of doping. Further research on this topic has been conducted using this R package and will be described elsewhere.

LIMITATIONS OF THE PACKAGE

As described above, the package allows a user to calculate the same ABPS values as available, for example, in the ADAMS database, thus providing a way to faithfully reproduce these results. However, as a consequence, the package inherits any potential shortcomings that the original ABPS algorithm may have. In particular, the original training dataset was based on athletes who received rhEPO injections (Sottas et al., 2006), and it may not be representative of other doping methods. For example, the ABPS will lose sensitivity with either volume expansion (IV or hyper hydration) or sub-micro dose EPO masking. As the authors of ABPS note (Sottas et al., 2006), the algorithm may need to be updated using new datasets when new blood doping methods are discovered.

The ABPS values are dependent on the blood analyzer used to generate the measures and are sensitive to bad pre-analytical and analytical conditions which have an impact on one or more of the seven hematological markers, such as variation in instrument

calibration; the protocols for collection, transport and analysis given in the WADA Athlete Biological Passport Operating Guidelines (WADA, 2018) should be used to guarantee a good interpretation of ABPS values. In addition, the distribution of the ABPS values, like the distribution of each of the seven hematological markers, is dependent on the population studied; different groups of athletes (for example, males vs. females) may display different distributions of ABPS values than the one obtained for the reference population on which the ABPS is based. For some usages of ABPS (for example, in a study of the prevalence of doping), it is thus advised to apply correction factors for the different populations. A future version of the package will include examples of such corrections, as described elsewhere (Robinson et al., 2018). It is worth noting that the ABPS algorithm restricts the possible input values to a range which depends on the values observed in the original reference dataset. If a value is outside this range, it will be modified to the minimum of maximum accepted value; for example, any HGB value lower than 12.9 or 18.2 will be modified to 12.9, respectively 18.2. The exact range for each parameter is indicated in the package; a warning will be printed if an input value is outside the allowed range.

AVAILABILITY OF THE PACKAGE

The R package is available either from CRAN (Comprehensive R Archive Network, https://cran.r-project.org/package=ABPS) or from a git server (https://gitlab.isb-sib.ch/BCF/ABPS). It is distributed under a free license, the GNU General Public License (version 2 or later).

In term of dependencies, the ABPS package requires only one package external to the base R system: kernlab, which is also available freely (licensed under the GNU GPL version 2) on CRAN.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

REFERENCES

CAS, (2012). Court of Arbitration for Sport. *Arbitration CAS 2012/A/2773.* Lausanne, November 2012. Available online at: https://jurisprudence.tas-cas. org/Shared%20Documents/2773.pdf

CAS, (2016), Court of Arbitration for Sport. *Arbitration CAS 2015/A/4006.* Lausanne, April 2016. Available online at: https://jurisprudence.tas-cas.org/ Shared%20Documents/4006.pdf

Gore, C. J., Parisotto, R., Ashenden, M. J., Stray-Gundersen, J., Sharpe, K., Hopkins, W., et al. (2013). Second-generation blood tests to detect erythropoietin abuse by athletes. *Haematologica* 88, 333–44.

Marrocco, C., Pallotta, V., D'alessandro, A., Alves, G., and Zolla, L. (2012). Red blood cell populations and membrane levels of peroxiredoxin 2 as candidate biomarkers to reveal blood doping. *Blood Transfus.* 10 (Suppl. 2):s71–7. doi: 10.2450/2012.011S

R Core Team (2017). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. Available online at:*https:// www.R-project.org/*

Robinson, N., Saugy, J., Schütz, F., Faiss, R., Baume, N., Giraud, S., et al.Saugy (2018). Worldwide distribution of blood values in elite track and field athletes: biomarkers of altered erythropoiesis. *Drug Test. Anal.* doi: 10.1002/dta.2526. [Epub ahead of print].

Sottas, P. E., Robinson, N., Fischetto, G., Dollé, G., Alonso, J. M., and Saugy, M. (2011b). Prevalence of blood doping in samples collected from elite track and field athletes. *Clin. Chem.* 57, 762–769. doi: 10.1373/clinchem.2010.156067

Sottas, P. E., Robinson, N., Giraud, S., and Saugy, M. (2006). Statistical classification of abnormal blood profiles in athletes. *Int. J. Biostat.* 2, 1557–4679. doi: 10.2202/1557-4679.1011

Sottas, P. E., Robinson, N., Rabin, O., and Saugy, M. (2011a). The athlete biological passport. *Clin. Chem.* 57, 969–76. doi: 10.1373/clinchem.2011.162271

Sottas, P. E., Robinson, N., and Saugy, M. (2010). The athlete's biological passport and indirect markers of blood doping. *Handb. Exp. Pharmacol.* 195, 305–26. doi: 10.1007/978-3-540-79088-4_14

WADA (2018). World Anti-Doping Agency. *Athlete Biological Passport Operating Guidelines.* Version 6.1. Available online at: https://www.wada-ama.org/ en/resources/athlete-biological-passport/athlete-biological-passport-abp-operating-guidelines

Zorzoli (2011). Zorzoli M. Biological passport parameters. *J. Hum. Sport Exerc.* 6:2. doi: 10.4100/jhse.2011.62.02