A third generation approach to detect erythropoietin abuse in athletes

Ken Sharpe
Michael J. Ashenden
Yorck O. Schumacher

Background and Objectives. Information derived from blood analyses can assist in the detection and/or deterrence of blood doping in sport. We investigated whether comparing an athlete's hematologic values against his or her own historical baseline rather than population-derived thresholds enhanced the ability to detect blood doping.

Design and Methods. We developed an approach whereby an athlete's true baseline value could be estimated with just one prior blood test. We also estimated a universal value for within-subject variability for key hematologic parameters using the highest value obtained among four separate cohorts of male athletes including 80 elite rowers, 124 endurance-trained or team-sport subjects, 288 professional football players and 630 athletes competing at national or international level. The (individual) baseline and (universal) variability were then incorporated so as to define expected thresholds for subsequent blood tests. The sensitivity of our approach was obtained by analyzing data from 49 recreational athletes administered either recombinant human erythropoietin (n=37) or placebo (n=12).

Results. We found that removing within-subject variability by comparing new results against an historical baseline heightened the capacity to detect blood doping. It was possible to delineate the longitudinal changes in either hemoglobin (Hb) or the OFF-hr model score (an algorithm using both Hb and percent reticulocytes) caused by recombinant human erythropoietin treatment from the natural biological fluctuations found in subjects treated with placebo.

Interpretation and Conclusions. Our objective data supported the intuitive belief that longitudinal monitoring of athletes' blood profiles will help detect blood doping. This information could be used to instigate target-testing of suspicious athletes, or even warrant the exclusion from competition of athletes with aberrant variations in key hematologic values.

Key words: hematologic passport, rHuEPO, athletes, blood doping.

Haematologica 2006; 91:356-363

From the Department of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia (KS); Science and Industry Against Blood doping (SIAB) research consortium, Gold Coast, Australia (MJA); Department of Sports Medicine, University of Freiburg, Freiburg, Germany (YOS)

Correspondence:
Ken Sharpe, Department of Mathematics and Statistics, University of Melbourne, Melbourne VIC 3010, Australia.
E-mail: k.sharpe@ms.unimelb.edu.au

In order to combat illegal blood doping, some sport federations exclude athletes from competition (*no start*) if they have abnormal hematologic values. This strategy is an important deterrent, even though thresholds must be set sufficiently high to accommodate inter-individual variation in the chosen hematologic parameters. In order to enhance the deterrent effect, it has been proposed that an athlete's blood values could be compared with his or her own historical values.[1] This so-called hematologic passport approach is intuitively simple; new values would be compared with the average of existing values obtained during previous blood tests, and new values that were substantially different from typical results for that athlete could lead to exclusion from competition. We propose that two criteria, hemoglobin concentration (Hb) and OFF-hr score,[2] merit consideration for inclusion in the passport. Hb is attractive because it reflects the oxygen carrying capacity of the blood, which is ultimately what all forms of blood doping seek to modify.[3] The hemoglobin assay is internationally standardized, and has better analytical characteristics than those of hematocrit.[4] The OFF-hr score, which is derived from an algorithm incorporating both Hb and reticulocyte percentage, is attractive due to its capacity to reveal recombinant human erythropoietin (rHuEPO) use many weeks after injections have stopped.[2] The International Cycling Federation (UCI) has already recognized the merit of the OFF-hr model and currently excludes athletes from competition who show aberrant OFF-hr score values. In addition to the prospect of excluding athletes from competition, as espoused by Malcovati *et al.*,[1] we believe that the passport offers an efficacious tool to identify hematologic fluctuations that are suspected to have been caused by prior blood doping, but which fall below the absolute level associated with a *no start*. This information could be utilized by antidoping agencies when compiling lists of athletes to be target-tested in the lead up to subsequent competitions. Because autologous and homologous transfusions induce hematologic perturbations in the recipient, the passport concept could also play a central role in screening for these practices. The aim of this paper was to evaluate the efficacy of using Hb and the OFF-hr score in a passport strategy. We first formulated a strategy to compare new blood results with historical values. Second, we quantified the expected

biological variation found in elite athletes for both Hb and the OFF-hr score. Finally we compared the sensitivity of our proposed passport approach with that of contemporary blood rules used by international sport federations.

Design and Methods

Subjects to quantify biological variation

Athletes from multiple countries who were active in a range of sporting disciplines participated voluntarily after giving their informed consent. In order to report statistically reliable sample sizes, only longitudinal data collected from male athletes are included in this paper. Four separate cohorts were studied, comprising predominantly rowers (French national rowing squad, France),[5] football players (professional and top level Italian clubs, Italy),[1]cycling/football/basketball/handball/tennis/triathlon (national and international level athletes, Germany) and state/national level athletes from multiple disciplines (International) (Table 1).[6]

Blood analysis

The rigor of analytical control able to be exercised during the phlebotomy procedures and sample analysis varied slightly between cohorts. For International data, all samples were collected according to a documented phlebotomy protocol then analyzed on ADVIA 120 hematology analyzers (Bayer Diagnostics, Tarrytown, NY, USA); these instruments were located in different countries[6] and calibrated according to the manufacturer's specifications. Italian data were collected in laboratories with quality control programs, however there were no specific controls exerted over the phlebotomy procedure.[1] The French samples were collected using medically-supervised phlebotomy procedures then measured on either an ADVIA 120 or a Sysmex Roche XE2100 (Sysmex, Kobe, Japan) that were both calibrated according to the manufacturers' specifications.[5] In Germany there was strict adherence to a standardized phlebotomy procedure and blood samples were analyzed on either an ADVIA 120 (to assay hemoglobin concentration) or a Sysmex Roche XE2100 (for reticulocyte values) that were calibrated according to manufacturers' specifications.

Passport rationale

Of fundamental importance to the passport concept is stipulating how many readings are required in order to establish the individual's baseline value against which new results are compared. There is no reason why a single previous reading cannot serve as an adequate baseline, provided that the formula used makes due allowance for the number of constituent readings. The simplest way to allow for different numbers of baseline observations is to use the formula: *variance between current and baseline readings*= $\sigma^2(1+1/n)$ where σ^2 is the variance between readings from the individual and n is the number of observations averaged to obtain the baseline reading. This formula assumes that, within an individual, all of the observations are independent of each other, which we propose can be satisfied if all samples used to derive the mean against which the current value is to be compared are taken at least one week apart (which does not preclude more frequent tests, as long as interim test results do not contribute to deriving a new mean score). Finally, σ^2 includes components for within-subject variation and between-days (including between-instruments where appropriate) variation.

Table 1. Estimates of variance for hemoglobin concentration (Hb) and OFF-hr score for longitudinal blood samples collected from each cohort. All subjects were male; the athletes from France were elite level rowers, those from Germany were predominantly cyclists, football players and rowers; those from Italy were top level football players; the International athletes were elite sportsmen from multiple disciplines. Variance components include between-subject, within-subject and between-days variation.

	France	*Germany*	*Italy*	*International*
Hb				
No. subjects	80	124	288	630
No. observations	409	603	841	1731
No. of days median	622	349	25	10
range[1]	(104-1502)	(31-662)	(6-1044)	(2-40)
Between-subjects	44.23	37.73	55.56	65.76
Within-subjects	19.27	26.95	17.35	15.41
Between-days[2]	7.10	12.91	5.21	6.03
Passport Hb3	26.37	39.86	22.56	21.43
OFF-hr				
No. subjects	80	122	288	630
No. observations	409	546	841	1731
No. of days median	622	332	25	10
range[1]	(104-1502)	(31-614)	(6-1044)	(2-40)
Between-subjects	84.66	66.19	83.67	124.41
Within-subjects	49.26	54.49	48.65	29.73
Between-days[2]	16.44	21.41	13.91	26.35
Passport OFF-hr[3]	65.70	75.90	62.56	56.08

[1]the number of days over which data were collected for individual subjects; median and (minimum–maximum); [2]between-days includes between-days, within-instruments and some between-instruments variation; [3]the values relevant for variation within subjects, including components for variation within- and between-instruments; these are the values that are appropriate for the passport. The 95% confidence interval for Passport Hb is 35.3–45.4; the 95% confidence interval for Passport OFF-hr is 66.6–87.3.

A z-score is derived using the formula:
$Hb_{z\text{-score}}=(Hb_{current}-Hb_{mean})/\sqrt{(\sigma^2\ (1+1/n))}$

where n=number of previous samples for the mean, and Hb_{mean} is the average of the Hb values of all samples taken prior to the current sample.

The variation of a current reading from the baseline reading can then be compared with a unit normal distribution in order to obtain a z-score, such that a z-score ≥+3.09 would be equivalent to a 1 in 1000 threshold and lead to a *no start* while a z-value ≥+2.33 (but < 3.09) or ≤-2.33 would be equivalent to a 1 in 50 threshold and lead to the athlete being targeted as *suspicious*.

Derivation of cut-off thresholds

In theory it would be possible to estimate a different value of σ^2 for each athlete, but this would make any application of the formula very complex and would likely require at least six samples for the baseline reading

and possibly considerably more. Therefore we opted to use the same σ^2 for all subjects. In order to estimate the variance σ^2, analysis of variance was used to obtain estimates of the variance components from each of the French, German, Italian and International cohorts. The variance components considered were between-subject, within-subject and between-days. The between-days component allows for within-instrument variation due to changes in operating conditions and recalibration, as well as differences between different instruments. For Hb there are well established methods for calibrating instruments, regardless of the type of instrument used, whereas for the the reticulocyte assay it is known that different brands of instrument (eg ADVIA and Sysmex) give substantially different readings. To allow for this all reticulocyte percentages used in a passport context would need to be calibrated using the so-called samplator method described by Ashenden *et al*.[7]

In order to facilitate an interpretation of the significance of a deviation of a new value from the designated baseline value, we opted to follow the notion of thresholds and the associated rate of false-positives which has been published concerning the OFF-hr blood model.[2] Two thresholds have been proposed: one is the 1 in 100 cut-off recommended as a threshold for suspicion deserving target-testing and/or medical follow-up, while the other, a 1 in 1000 cut-off, has been recommended for a *no start*. For both Hb and OFF-hr, a current value could be greater or less than the average of previous values. The most likely scenario for a drug cheat presenting at a competition venue would be for the athlete to have recently ceased rHuEPO treatment (so that all traces of rHuEPO would have left the system therefore ensuring a negative post-competition urine control sample). Following rHuEPO treatment, there is an increase in both Hb levels and OFF-hr score as compared with baseline values. Another scenario would be an athlete commencing rHuEPO treatment at the time of an out-of-competition test (or, less likely, at the time of the competition itself). This would present as a decrease in OFF-hr score (in tandem with an increase in Hb) as compared with baseline values collected during a period of no doping. On the other hand, if an athlete had been using some form of blood doping around the time the baseline samples were collected, but not around the time of the *current* sample, this might also show up as a decrease in Hb. A decrease in either parameter (Hb or OFF-hr) that exceeds a 1 in 100 cut-off could therefore be interpreted as a salient trigger for urine analysis and/or future target testing of the athlete. Using both the *1 in 100 increase* and the *1 in 100 decrease* cut-offs to target athletes the chance of completely *clean* athletes exceeding the cut-offs is 1 in 50.

Passport evaluation

To assess the utility of the passport concept to detect rHuEPO usage, and to ascertain what percentage of subjects using rHuEPO would exceed the relevant thresholds, we evaluated the fluctuations in Hb and OFF-hr model scores found to occur in subjects who were treated with rHuEPO during carefully controlled laboratory trials. The rHuEPO administration studies we evaluated were the basis for the second generation blood models that have been published previously.[2] We evaluated the Canberra and Oslo studies (the latter since the illustrated dosage regimen most closely represents the treatment protocols some contemporary athletes are suspected of using in order to avoid detection via blood tests, ie. low/maintenance doses of rHuEPO together with iron supplementation). It is noteworthy that the goal of the Oslo trial was to increase individual hematocrit levels by 10% and then to titer rHuEPO dosages in order to maintain these moderately elevated hematocrit values.

Since the longitudinal cohorts reported in this paper were formed of male subjects, we likewise restricted our evaluation to males who had participated in the rHuEPO administration trials. Briefly, the Canberra trial comprised 28 male recreational athletes who received 50 IU/kg rHuEPO three times a week for 3 weeks and then 20 IU/kg three times a week for the next 5 weeks, while eight males received placebo.[2] In the Oslo trial, nine males received ~40 IU/kg rHuEPO three times a week for 20 days and then ~18 IU/kg three times a week for the next 27 days, while four males received placebo.[2] To evaluate the sensitivity of the passport strategy, we used as a baseline the average of values collected before, and up to three days after, the first injection as an estimate of the athletes' natural values. We then evaluated data from the subsequent longitudinal visits which ended 21-24 days after the last injection.[2]

Results

Components of variation

The data available, which constituted multiple observations on many subjects, enabled the partitioning of the observed variation into between-subject, within-subject and between-days components of variance. Estimates of these components, obtained using standard analysis of variance methods, are given in Table 1. The different numbers of subjects and numbers of observations for Hb and OFF-hr for the German data are due to missing values of reticulocyte percentage for some samples. The variance required to evaluate changes in Hb (*Passport Hb*) or OFF-hr (*Passport OFF-hr*) scores using the passport concept is the sum of the within-subject and between-days variances, as shown in Table 1.

Unusually large passport variations of 39.86 for Hb and 75.90 for OFF-hr were found for the German data. This prompted us to carefully scrutinize the laboratory records, searching for evidence that the increased variability was due to the most obvious potential sources, such as a particular group of athletes, measurement error or data input errors. No explanation could be found for the atypical results which contrasted markedly with the remaining data sets.

Example of calculation

As an illustration of our passport concept, the following demonstrates the calculation of relevant change values for both Hb and the OFF-hr model. The example is a male athlete whose historical profile was: first blood sample Hb 148 g/L and percent reticulocytes of 1.10%,

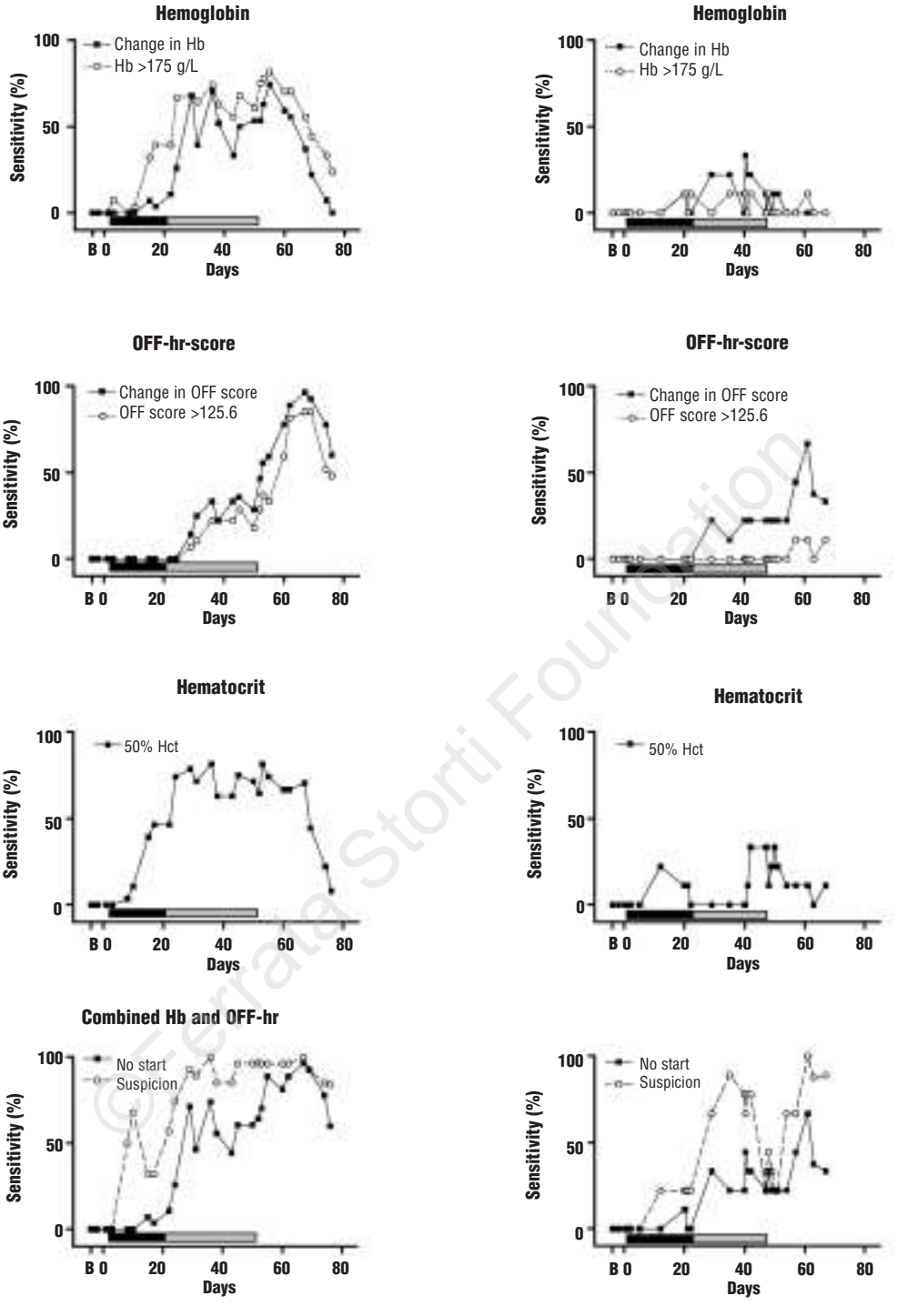

Figure 1. Time-course of the rate of detection of rHuEPO administration (sensitivity) of different detection strategies utilizing either Hb (top panels, either change in Hb exceeding a 1 in 1000 (false positive rate) cut-off or a single value exceeding 175 g/L), OFF-hr score (second tier panels, either change in OFF-hr score exceeding a 1 in 1000 cut-off or a single score exceeding 125.6), haematocrit (third tier panels, a single value exceeding 50%), or a combination of Hb and OFF-hr score (bottom panels, depicting either a 1 in 1000 *no start* cut-off or a 1 in 50 *suspicion* of doping cut-off). Two different rHuEPO administration protocols were evaluated: *Canberra* (leftside panels) denote sensitivity in n=28 male recreational athletes who received 50 IU/kg three times per week for 3 weeks (denoted by solid bar) then 20 IU/kg three times per week for the next 5 weeks (denoted by gray bar); *Oslo* (rightside panels) denote sensitivity in n=9 endurance-trained subjects who had competed previously at national level and who received ~40 IU/kg three times per week for 20 days (denoted by solid bar) then ~18 IU/kg three times per week for the next 27 days (denoted by gray bar). Sensitivity is expressed as the percent of subjects who exceeded the relevant cut-off.

corresponding to an OFF-hr model score of 85.07; second blood sample Hb 170 g/L and reticulocytes 1.13%, yielding an OFF-hr score of 106.22; third blood sample Hb 157 g/L and reticulocytes 0.87% yielding an OFF-hr score 101.04. For Hb we use σ^2=39.86 and for OFF-hr σ^2=75.90 (Table 1).

After the second blood sample is collected, a $Hb_{z\text{-score}}$ is derived using the formula:

$Hb_{z\text{-score}}=(170-148)/\sqrt{(39.86\ (1+1/1))}=2.464$

After the third blood sample is collected, a second $Hb_{z\text{-score}}$ is derived using the formula:

$Hb_{z\text{-score}}=(157-159.0)/\sqrt{(39.86\ (1+1/2))}=-0.259$

For the change in the OFF-hr model, after the second blood sample is collected, an $OFF\text{-}hr_{z\text{-score}}$ is derived using the formula:

$OFF\text{-}hr_{z\text{-score}}=(106.22-85.07)/\sqrt{(75.90\ (1+1/1))}=1.798$

After the third blood sample is collected, a second $OFF\text{-}hr_{z\text{-score}}$ is derived using the formula:

$OFF\text{-}hr_{z\text{-score}}=(101.04-95.65)/\sqrt{(75.90\ (1+1/2))}=0.505$

Sensitivity of the passport concept

Results from the Canberra trial illustrate that the sensitivity of the three approaches that reflect oxygen carrying capacity (change in Hb, Hb >175 g/L and hematocrit >50%) are most likely to detect athletes who are in the midst of receiving rHuEPO (see data points above the black and gray bars in Figure 1). The hematocrit exceeded 50% in 27 of the 28 subjects, resulting in 55.6% of readings (335/602) being above 50% hematocrit; four subjects exceeded 50% on between one and four occasions while 14 of the subjects exceeded it on 15 to 20 occasions. For controls 4.5% (9/199) of their readings exceeded a hematocrit of 50%. For Hb, 26 of the 28 subjects had values greater than 175 g/L such that 53.0% (319/602) of all readings were above this level; seven subjects exceeded the cut-off on between one and five occasions while 13 of the subjects exceeded it on 15 to 22 occasions. For controls Hb was above 175 g/L 4.0% (8/199) of readings. For change in Hb, all 28 subjects had at least three readings above the level of *suspicion*, which constituted 63.1% (380/602) of all data; all 28 subjects also exceeded the 1 in 1000 *no start* cut-off on at least one occasion, which constituted 35.7% of data readings (215/602). No values from the eight male controls exceeded either the *suspicion* or *no start* thresholds for change in Hb at any time point (0/168). The sharp decline in the sensitivity graphs for hematocrit, Hb and change in Hb after rHuEPO administration ceased illustrates that these parameters can lose their ability to detect rHuEPO use after treatment ceases. In contrast, the sensitivity of either change in OFF-hr score or OFF-hr >125.6 increased in the weeks after treatment ceased. For the OFF-hr model, 27 of the 28 subjects exceeded the 1 in 1000 cut-off value of 125.6; 13 subjects exceeded 125.6 on between one and five occasions while six of the subjects exceeded it on 10 to 16 occasions (a total of 173/602 or 28.7% of readings). No control subjects exceeded a value of 125.6. When evaluating the change in OFF-hr model score, we found that all 28 subjects had at least five readings that exceeded the *suspicion* cut-off (362/602 or 60.1% of all readings); moreover, all 28 subjects also exceeded the 1 in 1000 threshold at least twice (229/602 or 38.0% of all readings). For control subjects, 2.4% of all readings (4/168) exceeded the *suspicion* threshold, although no values exceeded the 1 in 1000 *no start* threshold. When utilizing changes in Hb and OFF-hr score in tandem, the sensitivity of the two approaches was modestly additive, inasmuch as heightened sensitivity was obtained both during and after rHuEPO treatment for the *no start* threshold (Figure 1, bottom panels). We did not alter thresholds when using the two parameters in tandem, therefore the overall rate of false-positives could be as high as 1:500. A total of 492 or 81.7% of all readings exceeded the *suspicion* threshold utilizing either the change in Hb or change in OFF-hr model, while 321 (53.3% of values) also exceeded the *no start* 1 in 1000 threshold for the combined parameters. No control subjects exceeded the *no start* threshold while four (of 168 or 2.4% of readings) exceeded the *suspicion* level for either the change in Hb or change in OFF-hr model. The capacity for surreptitious titration of rHuEPO dosages to confound existing blood models as

Table 2. Specificity of the Passport approach when applied to the four cohorts that were used to generate the model algorithms (Table 1), plus two groups of control subjects (Canberra, Oslo) who received placebo injections during double-blind rHuEPO administration trials (Figure 1). Evaluations were classified as 'suspicious' if passport values exceeded 1:100 cut-offs for either an increase or decrease in respective scores.

	Passport evaluations	*Passport evaluations identified as 'suspicious'*		*Passport evaluations exceeding 1:1000 'no start' threshold*	
Study		*Number (subjects)[1]*	*%*	*Number (subjects)[1]*	*%*
Hb					
France (n=80)	329	2 (2)	0.6	0	0
Germany (n=124)	479	13 (11)	2.7	0	0
Italy (n=288)	553	1 (1)	0.2	0	0
International (n=630)	1101	8 (8)	0.7	0	0
Canberra (n=8)	168	0	0	0	0
Oslo (n=4)	91	0	0	0	0
Total (n=1134)	*2721*	*24 (22)*	*0.9*	*0*	*0*
OFF-hr					
France (n=80)	329	6 (6)	1.8	0	0
Germany (n=122)	424	24 (22)	5.7	0	0
Italy (n=288)	553	11 (11)	2.0	4 (4)	0.7
International (n=630)	1101	25 (25)	2.3	2 (2)	0.2
Canberra (n=8)	168	4 (3)	2.4	0	0
Oslo (n=4)	91	4 (1)	4.4	0	0
Total (n=1132)	*2666*	*74 (68)*	*2.8*	*6.6*	*0.2*
Hb & OFF-hr					
France (n=80)	329	8 (8)	2.4	0	0
Germany (n=124)	479	35 (27)	7.3	0	0
Italy (n=288)	553	12 (12)	2.2	4	0.7
International (n=630)	1101	32 (32)	2.9	2	0.2
Canberra (n=8)	168	4 (3)	2.4	0	0
Oslo (n=4)	91	4 (1)	4.4	0	0
Total	*2721*	*95 (83)*	*3.5*	*6 (6)*	*0.2*

[1]*The number of subjects who exceeded the respective thresholds is given in brackets.*

well as absolute threshold limits for Hb and hematocrit can be easily discerned when comparing the sensitivity results between the Canberra and Oslo studies (Figure 1). Whereas around half of the subjects in the Canberra trial demonstrated Hb or hematocrit in excess of the relevant thresholds, the percentage of subjects in Oslo receiving a titered rHuEPO regimen who also exceeded the same nominal thresholds was substantially lower. For hematocrit, 13.1% (27/206) of values from treated subjects exceeded 50%, with one subject exceeding this value on 12 occasions and five subjects having one to six readings above 50%. There were 9/206 readings (4.4%) that exceeded a Hb value of 175 g/L, all of which came from one subject. None of the four male controls exceeded either the hematocrit or Hb threshold at any time point. The sensitivity of the absolute threshold of 125.6 for the OFF-hr model was similarly poor with only 1.5% of values (3/206, all from one subject) exceeding 125.6 (no controls had values above this threshold).

For change in Hb, all nine subjects had at least one reading in excess of the *suspicion* level. We found that 28.6% of values (59/206) exceeded the *suspicion* threshold while 8.3% of values (17/206) also exceeded the threshold associated with *no start* (0/91 from the controls exceeded the *suspicion* threshold). For change in OFF-hr all nine subjects had at least three readings that exceeded the *suspicion* threshold such that 34.0% of values (70/206) exceeded the threshold. Seven of the nine subjects also exceeded the *no start* threshold (19.9% or 41/206). Four of the 91 values from the control subjects (4.4%) exceeded the *suspicion* threshold, but none exceeded the *no start* threshold.

A combination of both change in Hb and change in OFF-hr score provided greater sensitivity than either approach alone. We found that 50.0% (103/206) of values exceeded the *suspicion* threshold, which corresponded with at least seven values from each of the nine subjects; 24.3% (50/206) of values (from eight of the nine subjects) also exceeded the 1 in 1000 threshold associated with a *no start*. Four values from the controls (4/91) exceeded the *suspicion* cut-off but no values exceeded the *no start* threshold.

With regard to the cohorts of athletes monitored during the data collection process and upon which our models have been based, we encountered several instances of changes in OFF-hr score or Hb that exceeded the *suspicion* and/or *no start* thresholds (Table 2). Passport evaluations were made by comparing each new value with the average of the previous values for the subject, so that the total number of evaluations for each group was the number of observations minus the number of subjects. A total of 0.9% of evaluations (24 of 2721), including those for the Canberra and Oslo controls, exceeded the *suspicion* threshold for change in Hb while 2.8% (74 of 2666) exceeded the *suspicion* threshold for change in OFF-hr; in each case the expected percentage was 2.0% (1 in 50) for *clean* athletes, corresponding to either a 1 in 100 increase or a 1 in 100 decrease. For change in OFF-hr, 0.2% (6 of 2666) evaluations also exceeded the 1 in 1000 *no start* threshold, compared with an expected value of 0.1%, while none of the changes in Hb exceeded the *no start* threshold.

Discussion

This study reinforces previous findings that a major component of the observed variation in the Hb values is between-subject variation. Malcovati *et al.*[1] conducted a nested analysis of variance on an undefined subset of 923 professional football players, and although the magnitude of the different variance components are difficult to extract from their paper, it was evident from the results that between-subject variation also represented the greatest source of variability in their data. Our current study, which includes a subgroup of the subjects analyzed by Malcovati, extends their findings on several levels. First, we included data collected from athletes participating in endurance disciplines such as rowers, road cyclists and triathletes. The results derived from these subjects were compatible with previous results. Second, we also evaluated the variance components of the OFF-hr model. In keeping with the pattern demonstrated by Hb, between-subject variance was also the major component of variance for the OFF-hr model score, lending scientific credence to the intuitive belief that a hematologic passport will hone and enhance antidoping strategies.

Establishing baseline values

An integral component of the passport concept is to establish a method of estimating an athlete's baseline hematologic values. Malcovati monitored 25 subjects over a 3-year period and, based on the resulting dispersion of within-subject coefficients of variation, concluded that it was necessary to obtain at least five determinations in order to reliably define subject-specific reference ranges.[1] For international sport federations wishing to implement a passport protocol, the collection of five longitudinal blood samples from the entire pool of candidate athletes imposes a significant logistical burden. For those sports that focus upon a *world cup* season and hold regular events at which a well-defined and relatively small (n<200) group of athletes compete, it is seemingly feasible to expeditiously collect five determinations from all competitors and to establish the subject-specific ranges proposed by Malcovati.[1] However for sports that may have in excess of 1000 participating athletes, or whose competitions are ubiquitous and truly international, the logistics of collecting and logging results in a central database pose a significant obstacle. It is vital to respect the financial and logistical constraints faced by antidoping bodies if an approach is to be considered for widespread implementation. Moreover, it is unclear how the five-test proposal would address the scenario in which a previously unheralded athlete bursts upon the international stage with immediate success but dubious blood values, since such an athlete would be outside the umbrella of a passport concept until such time that five collections had been made.

Our proposal to overcome this potential obstacle was to utilize an approach whereby a single previous value could serve as an estimate of the athlete's baseline value. This provides antidoping bodies with an expeditious method of implementing a passport concept. However

in the absence of subject-specific variations documented for each athlete, it dictates the adoption of a universal value representing the variance between readings for an individual. In the absence of clear evidence that within-subject variability was sufficiently different between athletes to confound our approach, we propose that the universal value for within-subject variation be equivalent to the highest value that we encountered in any of our subject cohorts (comprising 3584 observations from 1122 athletes for Hb and 3527 observations from 1120 athletes for OFF-hr). Two factors that can influence Hb and/or OFF-hr scores are whether or not samples are taken at altitude and, for OFF-hr, the type of instrument used to measure reticulocyte percentage. The latter can be allowed for by calibrating the instrument used, as detailed by Ashenden *et al*.[7] The potential effect of altitude is most readily allowed for by limiting values to be used in the passport to those obtained from samples collected at (or close to) sea level.

Defining cut-off thresholds

We incorporated the within-subject variability of the German cohort as our notional *universal* value, despite the fact that the variability in this group far exceeded that found in any other published data set. Rigorous pre-analytical control was in place during phlebotomy, negating the likelihood that obvious confounding influences such as posture or hydration led to the higher than expected variability. The technicians in charge of the blood analyzers could find no evidence of unusual control or calibration events during the period that data were collected. Statistical analyses were also conducted on various subsets of athletes to identify whether the variation was attributable to a particular group, but the variation was found to be reasonably uniform across different sports and different collection periods. No tenable explanation could be found to account for the greater within-subject variation found in the German cohort. Utilization of the German within-subject variability inevitably led to higher thresholds, which in turn meant that a greater percentage of drug users would show variability that was lower than the nominal thresholds. Indeed it could be argued that adopting this highest estimate is just as erroneous as adopting the smallest estimate we obtained. On the one hand this watered down the efficacy of our passport approach, and it is expected that the universal variance we have chosen will exceed achievable results when instruments and collection procedures are uniform and carefully monitored. However on balance we believe that it is important to instill some level of reassurance for clean athletes that they will not exceed the nominal thresholds, and utilizing the highest estimate provided even greater reassurance that the likelihood of a clean athlete exceeding the thresholds when demonstrating normal daily variations was correspondingly lower.

Exclusion from competition

Several sport federations already exclude athletes from competition if a blood check reveals unusually high Hb or hematocrit values (*no start*). For the International Cycling Federation (UCI), cyclists are barred for 14 days if their hematocrit exceeds a notional 50% threshold. For the International Ski Federation (FIS), skiers are excluded from competition if their Hb exceeds 175 g/L. Depending on which set of population-derived statistics are used, these thresholds overlap the population distribution values for healthy males such that approximately 1-5% of the male population would possess these values naturally. By contrast, it would be expected that only one sample per thousand from drug-free athletes would exceed the thresholds we have proposed for changes in both Hb and OFF-hr score. Passport evaluations from the various cohorts, including the Canberra and Oslo controls, resulted in none of the 2722 Hb evaluations, from 1134 athletes, exceeding the 1 in 1000 threshold (Table 2). However, for OFF-hr, six (0.2%) of the 2666 passport evaluations, from 1132 athletes, did exceed the 1 in 1000 threshold, which is somewhat higher than the 2.666 expected (the probability of observing at least six exceedances is 0.054). Possible reasons for this include actual use of blood doping by subjects within our cohorts (ie true-positives), hematologic abnormalities and sampling error.

A second tier to our passport concept is the ability to interpret blood results against multiple thresholds. For example, when adopting a threshold whereby 1 in 50 samples from clean athletes would be expected to exceed the score, the vast majority of subjects treated with rHuEPO exceeded this cut-off. Importantly, this trait held even for subjects who were treated with low doses of rHuEPO. Although the certainty of guilt associated with such values may be insufficient for a sport federation to feel justified in excluding the athlete from competition, such an aberrant value would seemingly warrant closer investigation. Utilizing the *suspicion* threshold as a trigger to initiate targeted out-of-competition testing may enable federations to rationalize their expenditure for out-of-competition testing programs, and simultaneously harness an increased deterrent effect by focusing attention on suspect athletes. For an athlete who provides a blood sample showing suspicious changes from historical values, a subsequent step may be to conduct an in-depth hematologic evaluation to identify any congenital or acquired hematologic disorders. We support the notion presented by Malcovati *et al*.[1] whereby a careful hematologic evaluation is undertaken, consisting of at least a blood cell count and iron status evaluation. The addition of more extensive testing proposed by Malcovati, including bilirubin, lactate dehydrogenase, serum iron, total iron binding capacity, serum ferritin, soluble transferrin receptor and serum erythropoietin levels, is in our opinion a logical and well-founded extension of this evaluation.

Further research

In the current study consideration has been limited to males as few females were included in three of the studies (seven in the German study, 29 in the French study and none in the Italian study). In all cases, estimates of variances that would have been used to assess changes in Hb or OFF-hr obtained from these females were less

than those obtained from males from the German study, which suggests that applying the cut-offs derived for male athletes to female athletes may be conservative. Until more data are obtained from female athletes it is recommended that the same cut-offs for changes in Hb and OFF-hr be used for both males and females.

Since altitude is known to affect our blood models, if sport federations wished to include blood values obtained from athletes whilst located at altitude, additional research would be required to ascertain the likely impact this would have on the passport concept.

In the current study we found that slightly higher percentages of athletes were identified as suspicious than predicted by our calculations when applied to the two newgroups of control athletes who were not included in the model derivation. Further insight regarding the actual percentage of athletes who exceed the suspicion threshold will be gleaned as federations begin screening their own athletes using our recommended models. Moreover, we expect that authorities would exercise their discretion when utilizing our models to identify athletes for target testing. Importantly, this issue would not impinge on the athletes eligibility to compete, since our models behaved as predicted with regard to the percentage of evaluations that exceeded the no-start threshold.

KS: analysis and interpretation of data, drafting and revision of manuscript, final approval of manuscript. MA: conception and design of study, drafting and revision of manuscript, final approval of manuscript. Both authors have agreed to the submission of the manuscript in the present form.

The authors wish to acknowledge the organizational and logistical support provided by Olaf Schumacher, Alain Lacoste, Michel Audran, Mario Cazzola, Luca Malcovati and Robin Parisotto that was fundamental to the collection of the hematologic data. We also wish to thank the subjects who volunteered for this study.

Conflicts of interest: none. Redundant publications: no substantial overlap with previous papers.

This study was principally supported by funds from the World Anti-Doping Agency (WADA). Data for the International cohort was collected during a study jointly funded by the Australian Government and the International Olympic Committee (IOC).

Manuscript received July 15, 2005. Accepted January 19, 2006. Prepublished online on February 17, 2006. PII: 03906078_9002.

References

1. Malcovati L, Pascutto C, Cazzola M. Hematologic passport for athletes competing in endurance sports: a feasibility study. Haematologica 2003; 88:570-81.
2. Gore CJ, Parisotto R, Ashenden MJ, Stray-Gundersen J, Sharpe K, Hopkins W, et al. Second-generation blood tests to detect erythropoietin abuse by athletes. Haematologica 2003;88:333-44.
3. Ashenden MJ. A strategy to deter blood doping in sport. Haematologica 2002;87:225-32.
4. d'Onofrio G. Hematology parameters: usefulness and limitations in monitoring red cell production. Bloodline Rev 2001;1:5-6.
5. Ashenden MJ, Lacost0e A, Orhant E, Audran M, Sharpe K. Longitudinal variation of hemoglobin and reticulocytes in elite rowers. Haematologica 2004;89:1403-4.
6. Sharpe K, Hopkins W, Emslie K, Howe C, Trout G, Kazlauskas R, et al. Development of reference ranges in elite athletes for markers of altered erythropoiesis. Haematologica 2002; 87:1248-57.
7. Ashenden MJ, Sharpe K, Damsgaard R, Jarvis L. Standardization of reticulocyte values in an antidoping context. Am J Clin Pathol 2004;121:816-25.