Second-generation blood tests to detect erythropoietin abuse by athletes

CHRISTOPHER J. GORE, ROBIN PARISOTTO, MICHAEL J. ASHENDEN, JIM STRAY-GUNDERSEN, KEN SHARPE, WILL HOPKINS, KERRY R. EMSLIE, CHRIS HOWE, GRAHAM J. TROUT, RYMANTAS KAZLAUSKAS, ALLAN G. HAHN

Background and Objectives. We previously developed blood tests that were introduced at the Sydney 2000 Olympic Games to identify athletes injecting recombinant human erythropoietin (rHuEPO). The aim of this study was to re-analyze our existing database to develop models with heightened sensitivity, using wherever possible blood parameters measurable with appropriate standards of analytical performance.

Design and Methods. The principal database for this study was derived from a double-blind trial in which 57 recreational athletes were administered either rHuEPO or placebo. Standard discriminant analysis was used to derive two ON models (ON-hes and ON-he) and two OFF models (OFF-hr and OFF-hre) sensitive to accelerated and decelerated erythropoiesis respectively, utilising concentrations of hemoglobin (h), erythropoietin (e) and serum transferrin receptor (s), as well as percent reticulocytes (r). The ability of our models to detect rHuEPO administration was assessed by comparing model scores of subjects in the administration trial with the model scores of 1152 elite athletes from 12 countries.

Results. The ability of the new models to detect rHuEPO administration was generally higher than that of our previous models, particularly during phases when low doses of rHuEPO were used, and after injections had ceased.

Interpretation and Conclusions. The increased stability of the new blood parameters facilitates transport of samples to central laboratories, and the heightened sensitivity of the new models makes them better than existing models for federations wishing to screen samples for urine testing and to identify and target suspect athletes for out-of-competition testing. However procedures should be incorporated that respect an elevated model score caused by genetic, health or environmental circumstances.

Key words: rHuEPO, athletes, blood tests, doping, erythropoiesis

Haematologica 2003; 88:333-344
http://www.haematologica.org/2003_03/88333.htm

From the Dept. of Physiology, Australian Institute of Sport (AIS), Canberra, Australia (CJG, RP, AGH), Science and Industry Against Blood doping (SIAB) research consortium, Gold Coast, Australia (MJA), Norway University of Physical Education and Sport, Oslo, Norway (JS-G), Dept. of Mathematics and Statistics, The University of Melbourne, Melbourne. Australia (KS), Sports Studies, Auckland University of Technology, Auckland, New Zealand (WH) and Australian Sports Drug Testing Laboratory (ASDTL), Australian Government Analytical Laboratories Sydney, Australia (KRE, CH, GJT, RK).

Correspondence: Christopher J. Gore, Australian Institute of Sport, Adelaide SA 5022, Australia. E-mail: chris.gore@ausport.gov.au

The realization of a urine-based test[1] that could identify athletes using the previously undetectable drug recombinant human erythropoietin (rHuEPO) was a watershed in antidoping research into detection of peptide hormones. Despite the deterrent effect of this test upon athletes tempted to use rHuEPO, several of its characteristics have attracted comment. First, the electrophoretic technique used to separate endogenous and exogenous erythropoietin (based on differential isoelectric profiles) is currently expensive and time-consuming. Second, the test lacks sensitivity if a urine sample is collected more than 3-4 days after the last injection of rHuEPO (because of the rapid clearance of rHuEPO). This implies that if a urine sample is collected from an athlete who had either not injected rHuEPO, or had done so more than 3-4 days previously, analysis of their urine sample would be futile.

Fortuitously, the expense and lack of retrospectivity of the urine test are complemented by the characteristics of previously published blood tests developed to identify athletes using rHuEPO.[2] Hematologic parameters sensitive to the rate of red cell production are disturbed for up to four weeks after treatment with rHuEPO, and the assays to measure these parameters are relatively inexpensive. It is, therefore, not surprising that sporting federations and antidoping authorities have expressed interest in the potential for information gleaned from blood tests to identify those urine samples with the highest likelihood of containing rHuEPO.

Quite separate from the financial/practical rationale for utilizing the physiologic information derived from blood analyses, is the need to counter the self-evident threat to sporting integrity posed by advances in pharmacology and drug therapy. The biotechnology race to mimic endogenous proteins and hormones threatens to out-reach the antidoping community's ability to delineate between exogenous and endogenous products. One example is the development of Dynepo, which is produced in a human cell line[3] and therefore, in contrast to rHuEPO and Aranesp,[4] may possess human glycosylation. It appears inevitable that in the future it will not be tenable to base antidoping deterrents solely upon the pharmacologic approach. Therefore it seems prudent to devote resources pro-actively to foster and refine alternative approaches to identify doping practices. Blood-based algorithms, capable of identifying athletes who have or are using an erythropoietic stimulant, are one promising alternative. Like most other antidoping researchers who have sought a means of identifying athletes abusing rHuEPO[5-8] we used relatively moderate

amounts (50-200 IU/kg^{-1}, approximately three times per week) for relatively short periods (10-30 days) to quantify the hematologic disturbance caused by rHuEPO injections.[2] Total red cell mass increases rapidly with such a regimen, and continued dosing of healthy athletes at this rate is unsustainable without risking thrombolytic events[9] or even death.[10] Therefore we anticipate that athletes would tend to use a lower dose of rHuEPO, but continuously. Because this scenario is subtly different from the circumstances in which we originally conceived blood tests for rHuEPO use, it presented an opportunity to revisit the inclusion and emphasis of individual hematologic parameters in our algorithms. We also wished to incorporate suggestions and advice we have received following application of our previous models in specific antidoping situations.

The aims of this research were three-fold and complementary. First, to develop models specifically tailored to serve a screening role in the selection of urine samples to be analyzed for the presence of rHuEPO. Second, to develop models that possess enhanced sensitivity to the marginally increased rates of erythropoiesis anticipated to occur in an athlete titrating their rHuEPO dose to maintain modestly elevated hemoglobin (Hb) levels. We were particularly conscious of the potential benefits conferred by models that highlighted blood profiles consistent with the cessation of rHuEPO use. Finally, our original models included several volume-dependent red cell characteristics, such as hematocrit (Hct), reticulocyte hematocrit and percent macrocytes. Because red blood cells can swell during storage, we sought to incorporate volume-independent parameters that would allow for increased storage/transport time between sample collection and laboratory analysis.

Design and Methods

Subjects

Two cohorts of subjects were studied. One cohort was formed of recreational athletes resident in Canberra Australia (Canberra, n = 57; 23 women, 34 men), the other cohort comprised were residents of Oslo, Norway (Oslo, n = 22; 9 women, 13 men). No Canberra subject was a member of a national sporting squad, but all had been in regular training during the year preceding the study (range; 1.5–20 hr/wk^{-1}) and continued to train throughout. All but one of the Oslo subjects had previously competed at national level in cross-country skiing, and the sole exception had been nationally competitive in another sport. Six had represented Norway in international events and one was a former Olympic medallist. However, at the time of the study, none of the Oslo subjects was involved in elite-level sport. All subjects provided signed informed consent, following procedures that had been approved by the Ethics Committee of the Australian Institute of Sport (Canberra group) and the Norwegian University of Sport and Physical Education (Oslo group).

Study design

These two, complementary, double-blind studies comprised 7-8 weeks of rHuEPO (or placebo) administration and an ~3 wk *wash-out* during which time subjects were monitored but injections had ceased (Table 1). After medical screening for illness, injury, and high blood pressure, the Canberra subjects were randomly assigned to one of three subgroups: group A (n = 11M/ 5F, rHuEPO + intravenous (IV) iron), group B (n = 16M/14F, rHuEPO + oral tablet iron), group C (n = 7M/4F, placebo). The results from groups A and B were subsequently combined into a single group for the purpose of this study. The Oslo subjects were likewise screened and randomly assigned to one of two groups: group D (n = 9M/6F, rHuEPO + oral liquid iron) group E, (n = 4M/3F, placebo). Therefore groups A, B and D constitute treatment or *rHuEPO* groups, whilst groups C and E were controls.

In Canberra, rHuEPO (EPREX 4000, Janssen-Cilag, Australia) or saline (NaCl 0.9% BP) injections were given subcutaneously (buttock) 3 times per week for 8 weeks; the first 3 weeks at a dosage of 50 IU/kg^{-1} and the next 5 wk at 20 IU/kg^{-1}. In Oslo, the injections were 3 times per week for 7 weeks, and the mean (±SD) dosage of rHuEPO was 40±4 IU/kg^{-1} for 20 days and then 18±3 IU/kg^{-1} for a further 27 days. The exact doses were varied to increase Hct and then maintain it ~10% above the individual's initial level. Iron infusions in Canberra were 100 mg IV (Ferrum H; iron hydroxide polymaltose, Dextrin, Sigma Company Ltd.) delivered fortnightly over 4 hr via an antecubital vein. Group B, took iron tablets (Ferrogradumet, Abbott, Australia) daily which provided ~105 mg of elemental iron derived from 350 mg of dried ferrous sulphate. Group D took daily liquid iron supplements (Neo-Fer, Nycomed, 9 mg/mL^{-1}) throughout the study. Dosages were individualized on the basis of measured serum ferritin concentrations at baseline, week 4 and week 8. The mean (±SD) iron dosages for the rHuEPO group were 146±44, 159±64 and 135±52 mg for weeks 1-4, weeks 5-8 and weeks 9-11, respectively. The corresponding figures for the placebo group were 139±44, 189±78 and 147±45 mg.

The compliance rates for the subjects injected with rHuEPO or saline were 98 and 100% in Canberra and Oslo, respectively. Twenty-seven of a possible total of 1425 injections (57 subjects × 25 injections) were missed in Canberra mostly as a consequence of an excessive Hct (in accordance with health and safety guidelines imposed by the

Table 1. Outline of the study design to examine the effects of low-dose (~20 IU/kg-1) administration of rHuEPO on blood parameters. Subjects received three rHuEPO injections per week for 8 weeks in the Canberra trial and 7 weeks in the Oslo trial. In Oslo, the exact doses were varied to increase Hct and then maintain it ~10% above the individual's initial level.

	Time/phase			
	~1-2 weeks	*~3 weeks*	*~4-5 weeks*	*~3 weeks*
Overview	Subject screening	Acceleration phase (to rapidly elevate hematocrit)	Maintenance phase (to sustain but not increase hematocrit further)	Follow-up (washout phase after cessation of injections)
	Canberra Trial			
	Weeks 1-2	*Weeks 3-5*	*Weeks 6-10*	*Weeks 11-14*
rHuEPO injection		Three per week at 50 IU/kg^{-1} (day 0, 2, 4, 7, 9, 11, 14, 16, 18 & 21)	Three per week at 20 IU/kg^{-1} (day 23, 25, 28, 30, 32, 35, 37, 39, 42, 44, 46, 49, 51, 53 & 56)	
Blood collection	Two baseline samples. (day -21 and day 0* relative to the first rHuEPO injection)	Twice per week: (day 1, 3, 8, 10, 15, 17 & 22)	Twice per week: (day 24, 29, 31, 36, 38, 43, 45, 50 & 52)	Twice per week: (day 1, 3, 8, 10, 15, 17, 22 & 24 after last rHuEPO injection)
Blood sample number	1&2	3-9	10-18	19-26
	Oslo Trial			
	Week 1	*Weeks 2-4*	*Weeks 5-8*	Weeks 9-11
rHuEPO injection		Three per week at ~40 IU/kg^{-1}: (day 0, 2, 5, 7, 9, 12, 14, 17 & 20)	Three per week at ~20 IU/kg^{-1}: (day 22, 24, 27, 29, 31, 33, 35, 37, 40, 42, 44 and 47)	
Blood collection	Two baseline samples. Day -7 and day 0* (relative to first rHuEPO injection).	Day 0.25, 1, 2*, 5*, 12* 20*, 20.25 and 21	Day, 22*, 29*, 35*, 40*, 40.25, 41, 42* and 47*	Day 0.25, 1, 2, 3, 4, 7, 10, 14, 17 and 21 (days after last rHuEPO injection)
Blood sample number	1&2	3-10	11-18	19-28

**When blood collection and rHuEPO injections occurred on the same day, the blood collection preceded the injection.*

Ethics Committee, saline injections were given whenever the Hct for the preceding blood sample was ≥ 0.55). A total of 13 of the 46 subjects on rHuEPO in Canberra were injected with saline instead of rHuEPO on at least one occasion, and two subjects had more than 5 saline injections. One subject developed elevated blood pressure after 3 weeks of rHuEPO administration and was injected with saline only for the remainder of the study. No Hct readings of 0.55 or above were recorded at any stage during the Oslo study, and consequently members of Group D were injected with rHuEPO on all scheduled occasions.

Blood analysis

The techniques for blood collection and analysis were the same as those we have described previously.[2] Briefly, whole blood samples (analyzed within 8 hr of collection) provided erythrocyte and reticulocyte parameters measured with an ADVIA120 Hematology Analyzer (Bayer Diagnostics, Tarrytown, NY, USA). One ADVIA was located in Australia at the Australian Institute of Sport and the second was located in Oslo, Norway. Frozen serum aliquots from Oslo were freighted on dry ice to Australia, where measurements were made of serum concentrations of erythropoietin (EPO) and

soluble transferrin receptor (sTfr) for both the Canberra and Oslo samples.

Each ADVIA was calibrated against appropriate reference materials, and controlled daily using Bayer ADVIA *TEST*point Haematology Low, Normal and High controls (no High control for Oslo) and Bayer ADVIA *TEST*point Reticulocyte Low and High controls. The average coefficient of variation (CV) for the parameters used in the models were as follows: Canberra ADVIA Hb 0.8%, 0.8% and 0.7% for Low, Normal and High controls respectively; % reticulocytes 14.9% and 3.6% for Low and High control respectively. For the Oslo ADVIA, CVs were: Hb 2.7% for Low and 0.9% for Normal controls (no High control available); % reticulocytes 16.0% and 5.4% for Low and High controls, respectively.

The EPO concentrations were determined using an automated solid-phase, sequential chemiluminescent Immulite EPO assay (Diagnostic Products Corporation, Los Angeles, CA, USA) and sTfr concentrations were measured by means of an automated immunonephelometric assay (Dade Behring GmbH, Marburg, Germany). The automated immunoassays for EPO and sTfr were controlled using three and two levels of controls, respectively. Using three levels of EPO controls (mean 15.2, 30.4 and 62.3 mU/mL^{-1}), the within-assay CVs were 4.7, 7.1 and 5.1%, and between-assay CVs were 7.3, 7.2 and 9.5%, respectively. Using two levels of sTfr controls (mean 0.63 and 1.45 mg/L^{-1}), the within-assay CVs were <1.0% and 2.2%, and between-assay CVs were 3.4 and 2.8%, respectively.

Statistical analyses

Model derivation. We developed two ON models and two OFF models[2] using data from the Canberra administration trial. The models use different subsets of the blood parameters Hb, percent reticulocytes, EPO and sTfr. The four models are referred to as ON-he, ON-hes, OFF-hr, and OFF-hre; these represent linear combinations of functions of the blood parameters hemoglobin concentration (h), serum EPO concentration (e), percent reticulocytes (r), and serum transferrin receptor concentration (s).

Standard discriminant analysis was used to derive the models. To obtain model scores with distributions as close to normal as possible, the values of percent reticulocytes were square-root-transformed and those of EPO and sTfr were log-transformed before analysis.

To allow for sampling variation between subjects in the rHuEPO and control groups, the values of each variable were adjusted by a constant amount so that the means of the rHuEPO and control groups were equal over visits 1 and 2 (the baseline blood samples). Separate adjustments were made for males and females. Discriminant analysis was then used on data from multiple blood samples for each subject. Data from visits 15 to 18 (the phase of low-dose administration) were used to derive the ON models, and data from visits 19-24 to derive the OFF models (Table 1). For ease of interpretation, the coefficients of the models were scaled to make the coefficient of Hb unity. In developing the models, data from males and females were pooled, and an allowance made for gender, so that the coefficients of the variables would be the same for males and females.

Cut-offs and false-positive rates for model scores. Having derived each model, we estimated cut-off scores corresponding to various *false-positive* rates by analyzing the model scores of the 1152 athletes in a previous study,[11] under the assumption that none of these athletes was a current or recent user of rHuEPO. We calculated each athlete's model score for each of their visits (up to three) to the laboratory, then analyzed the scores using the mixed modeling procedure (Proc Mixed) in the Statistical Analysis System (SAS Version 8.02, SAS Institute, Cary, NC, USA) to estimate values for the same fixed and random effects as for the old ON- and OFF-model scores reported in the previous study. The fixed effects were ethnicity, age, sport, altitude, visit, time since last exercise, and time of day of the visit; the random effects allow for within- and between-athlete and within- and between-ADVIA variations.[11] Scores from the two OFF models were power transformed (by 1.8 for OFF-hre and 1.5 for OFF-hr) prior to analysis, to deal with substantial departure from normality that was evident in the lower tail of the distribution. Cut-offs were subsequently back transformed to the original scale.

We derived cut-offs first for the most common endurance athletes (Caucasian athletes of age 19-24 years, who were tested at <610 m above sea level, and whose blood sample was collected on the first laboratory visit between 07:30AM and 4:00PM at least 12 hours after exercise). Endurance athletes were chosen for the reference group as scores on all models were, on average, higher for these athletes, and it is expected that endurance athletes are the most likely to seek a competitive advantage from use of rHuEPO. Cut-offs for this group of endurance athletes (hereafter referred to as the *typical* group) were estimated by combining the mean, standard deviation, and degrees of freedom with the value from the t distribution that gave a one-tailed probability equivalent to each of 13 arbitrarily chosen false-positive rates between 1 in 10 and 1 in 10,000. These theoretical false-positive rates were then plotted on a log scale against cut-off, and the points connected with a smooth curve (Figure 1).

To check on the correspondence between theoretical and observed false-positive rates for a given cut-off score, we tallied the model score of each

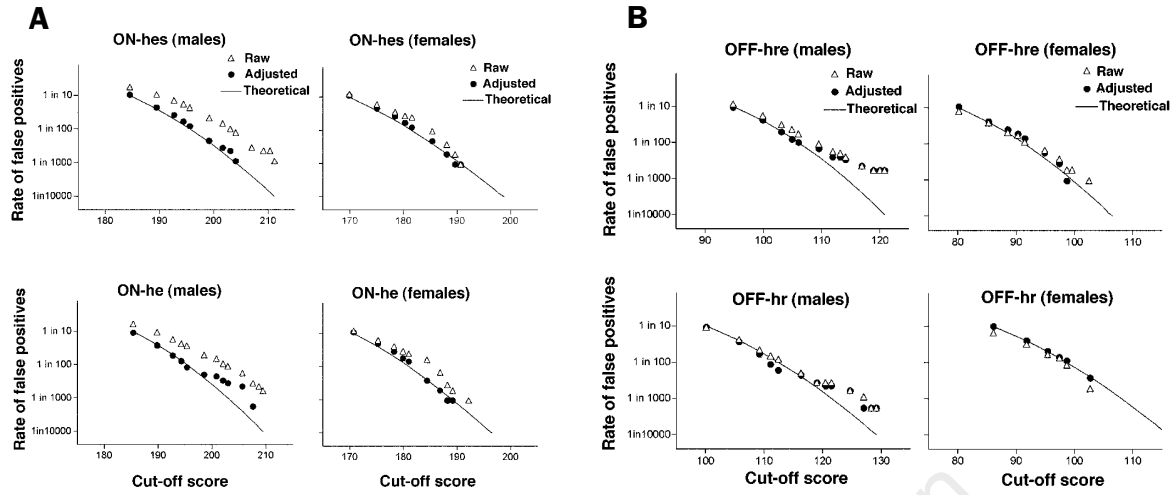

Figure 1. The relationship between false-positive rate (plotted on a log scale) and cut-off score for ON models (Figure 1A) and OFF models (Figure 1B). The solid line is the theoretical false-positive rate for typical endurance athletes at sea level, based on assumptions of normality of the distribution of model scores and applicability of the fixed- and random-effects models in the study of 1,152 elite athletes.[11] Raw false-positive rate (open triangles) is the rate observed when model scores were calculated for those athletes without adjustment. Adjusted false-positive rate (filled circles) is the rate when scores were adjusted to those of typical endurance athletes at sea level using appropriate levels of the fixed effects. The fixed effects, as described previously (See *Design and Methods*), are ethnicity, age, sport, altitude, visit number, time since exercise and time of day.

typical athlete as a false positive if it exceeded the cut-off; for athletes on other levels of the fixed effects, we subtracted the mean value of the appropriate levels of the fixed effects from each athlete's model score before tallying the score. The resulting tally, expressed as a percent, was defined as the observed false-positive rate and was plotted on Figure 1.

This definition is based on the assumption that the standard deviation and degrees of freedom for the distribution of athletes on any combination of levels of the fixed effects were the same as for athletes in the typical group; inspection of the standard deviations and degrees of freedom for different levels of the fixed effects showed that this assumption would result in negligible error in estimation of cut-offs or false-positive rates.

To illustrate the impact of the fixed effects on false-positive rates, we also calculated the observed false-positive rate when we did not adjust athletes' scores for the fixed effects (referred to hereafter as *raw* model scores). These points are also plotted on Figure 1.

In addition to the cut-off scores for the typical (endurance) athlete, we also calculated the cut-off scores that would produce the same false-positive rates for an athlete on those levels of the fixed effects that give the greatest positive model score (referred to hereafter as *worst-case* athletes). The levels of these fixed effects varied somewhat between models and sexes, but in all cases being tested at altitude had a positive effect (levels of the fixed effects for worst-case athletes and the effect on model scores are available in the on-line appendix). Collecting a blood sample whilst the athlete was at altitude had the greatest positive effect on all scores, however because the collection (or not) of samples at altitude is at the discretion of authorities, we also calculated the cut-off score that would produce the same false-positive rate for the worst-case athlete at sea level.

Sensitivity. Sensitivities of the models were assessed by comparing model scores of subjects in the rHuEPO groups of the administration trials with the cut-offs determined as above. The model scores were first adjusted by a constant amount to make the mean at baseline equal to the mean score for the typical group. This adjustment allowed for any systematic difference between the subjects in the rHuEPO groups and the typical group. Separate adjustments were made for males and females from each of the rHuEPO groups. The sensitivity of the model arising from scores for a given visit to the laboratory was expressed as the percent of subjects whose model scores exceeded the given cut-off.

To compare the sensitivity of the models during different phases of rHuEPO use, we plotted these sensitivities against day of visit for cut-offs associated with false-positive rates of 1 in 100 for the ON models and 1 in 1,000 for the OFF models (Figure 2). To compare the new models with those derived previously, sensitivities for the *old* ON and OFF models

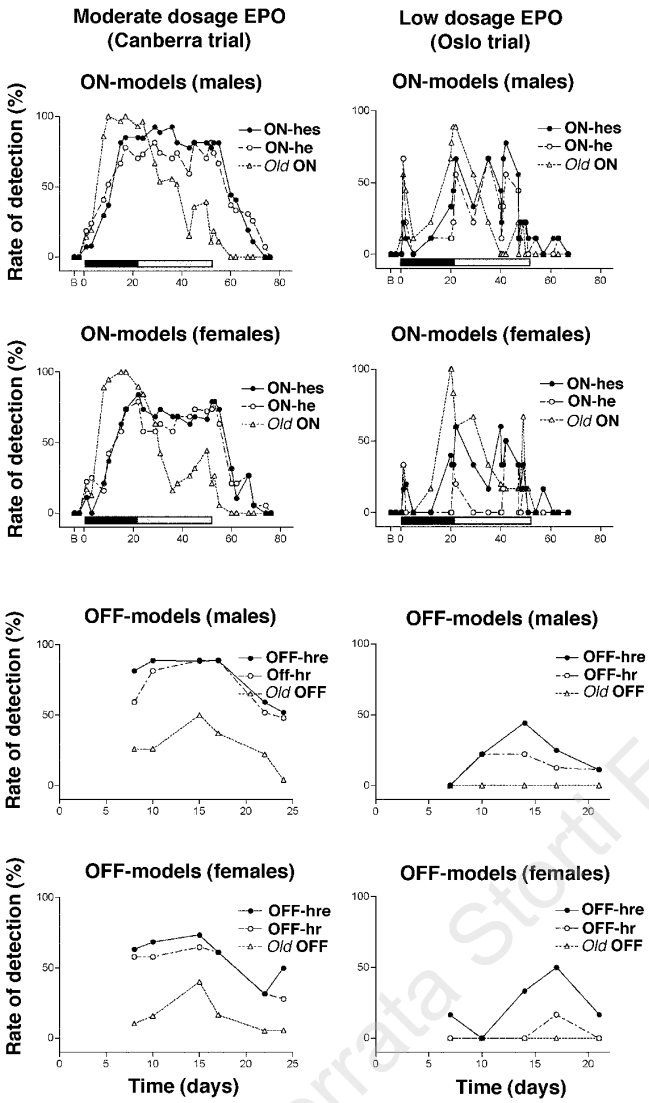

Figure 2. Time-course of the rate of detection of rHuEPO administration (sensitivity) for the four new models (ON-hes, ON-he, OFF-hre and OFF-hr) and the two old models ('old' ON and 'old' OFF)[2]. Sensitivities were obtained following the first injection (ON models) and after the last injection (OFF models) of rHuEPO. In each trial, athletes' scores were adjusted by a constant amount to make their means at baseline the same as those of worst-case endurance athletes tested at sea level. Sensitivity is expressed as the percent of athletes whose adjusted scores exceeded the cut-off scores corresponding to a false-positive rate of 1 in 100 (ON models) and 1 in 1000 (OFF models). The solid bar depicts acceleration phase, and the grey bar depicts maintenance phase, of rHuEPO injections.

derived in a similar manner were also plotted.

To demonstrate the effect of false-positive rate on sensitivity, we first generated cut-off scores corresponding to a range of false-positive rates for a particular group of athletes. We chose endurance athletes tested at sea-level with characteristics that, on average, produce the highest model scores (worst-case athletes). These cut-offs were then used to estimate rates of detection of rHuEPO administration for athletes during appropriate phases in the Canberra and Oslo studies. For the ON models, these phases corresponded with administration of a low dose of rHuEPO (days 29-52 in the Canberra study; days 29-47 in the Oslo study). For the OFF models, we nominated periods 8-22 days (Canberra) and 7-21 days (Oslo) following the cessation of (low dose) rHuEPO injections. Estimates of sensitivities were then obtained as the proportion of all model scores from all athletes for the period that exceeded the cut-off. For each model we then plotted these sensitivities against the false-positive rate (Figures 3A and 3B).

Results

The models we chose for analysis of sensitivity were as follows:

$$\text{ON-he} = \text{Hb} + 9.74\ln(\text{EPO});$$
$$\text{ON-hes} = \text{Hb} + 6.62\ln(\text{EPO}) + 19.4\ln(\text{sTfr});$$
$$\text{OFF-hr} = \text{Hb} - 60\sqrt{(\text{Ret})};$$
$$\text{OFF-hre} = \text{Hb} - 50\sqrt{(\text{Ret})} - 7\ln(\text{EPO}).$$

Abbreviations (and units): ln, natural logarithm; Hb, hemoglobin concentration (g/L); reticulocytes (%); EPO, erythropoietin concentration (mU/mL^{-1}); sTfr, serum transferrin receptor (mg/L^{-1}).

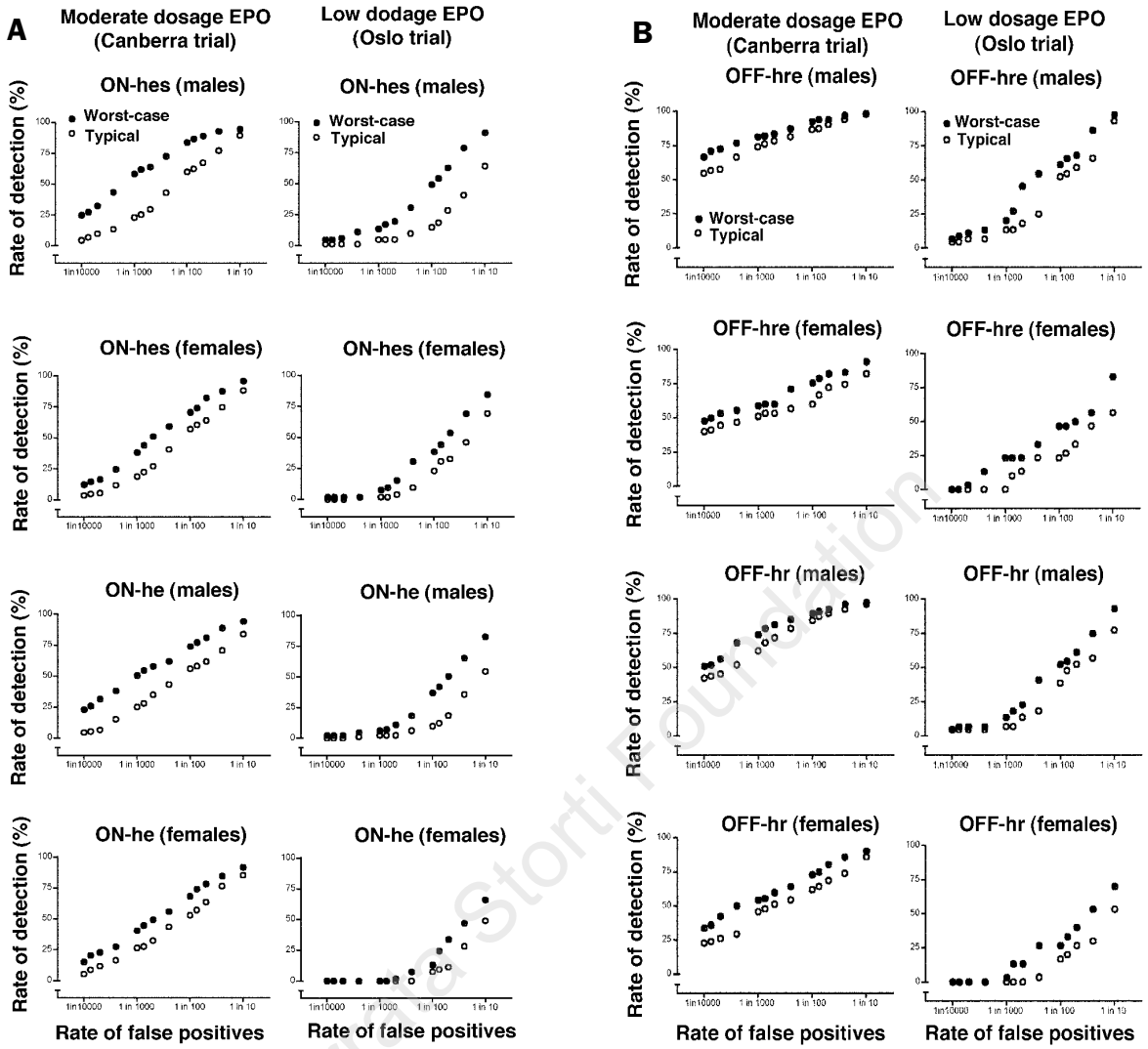

Figure 3. The effect of choosing various false-positive rates on the rate of detection of rHuEPO administration (sensitivity) with the new ON (Figure 3A) and OFF (Figure 3B) models. The left panel depicts sensitivity associated with moderate doses of rHuEPO (Canberra trial), the right panel depicts sensitivity associated with low doses of rHuEPO (Oslo trial). For both trials, subject's scores were adjusted by a constant amount to make their means at baseline the same as those of typical or worst-case endurance athletes tested at sea level. The sensitivities represent the percentage of worst-case (filled circles) or typical (open circles) subjects administered low or moderate doses of rHuEPO who exceeded the cut-off scores for worst-case endurance athletes tested at sea level corresponding to each of the thirteen false-positive rates.

The relationships between false-positive rates and cut-off scores (based on 1152 elite athletes)[11] are shown in Figure 1 for all the models. In some cases false-positive rates for some of the highest cut-off scores do not appear in the figures because the rates fell to zero and so cannot be plotted on a log scale. For the ON models, raw model scores (not adjusted for fixed effects) produced false-positive rates markedly highly than the theoretical rate over the entire range of cut-off scores for both sexes. By comparison, the raw model scores for the OFF models produced false-positive rates that were close to the theoretical values (except for the OFF-hre in males).

When each athlete's raw score was adjusted for the levels of the fixed effects characterizing that athlete, the resulting false positive rate for females followed the theoretical rate very closely. For males the false-positive rates for the adjusted model scores followed the theoretical rate closely (or was zero) for ON-hes, but showed systematic deviations towards rates higher than predicted for rates

Table 2. Cut-off scores corresponding to selected false-positive rates for typical endurance athletes (those with the most frequent characteristics) tested at sea level (<610 m), worst-case endurance athletes (those with characteristics producing the highest model scores) tested at sea level, and worst-case endurance athletes tested at altitude (>610 m), for the two ON and two OFF models. Cut-offs and false-positive rates for typical athletes at sea level were used to plot the theoretical curves shown in Figure 1 and comparisons between the old and new models shown in Figure 2. Cut-offs for worst-case athletes at sea level were used to derive the rates of detection shown in Figure 3.

	Males			*Females*		
	Typical at sea level	*Worst-case at sea level*	*Worst-case at altitude*	*Typical at sea level*	*Worst-case at sea level*	*Worst-case at altitude*
ON-hes						
1 in 10	184.5	192.2	209.4	169.9	174.6	187.6
1 in 100	195.6	203.3	220.5	181.6	186.3	199.3
1 in 1000	204.0	211.7	228.9	190.7	195.4	208.4
1 in 10000	211.2	218.9	236.1	198.8	203.5	216.5
ON-he						
1 in 10	185.4	191.4	207.1	170.7	175.0	187.6
1 in 100	195.3	201.3	217.1	181.0	185.3	197.9
1 in 1000	202.9	208.9	224.7	189.2	193.5	206.0
1 in 10000	209.5	215.5	231.2	196.5	200.8	213.3
OFF-hre						
1 in 10	94.8	99.6	108.1	80.2	86.8	94.1
1 in 100	106.0	110.5	118.3	91.5	97.4	104.1
1 in 1000	114.1	118.3	125.7	99.6	105.2	111.6
1 in 10000	120.8	124.8	132.0	106.5	111.8	117.9
OFF-hr						
1 in 10	100.1	104.6	113.7	86.1	92.2	99.9
1 in 100	112.4	116.7	125.3	98.7	104.4	111.7
1 in 1000	121.5	125.6	134.0	108.1	113.5	120.5
1 in 10000	129.2	133.2	141.3	116.1	121.4	128.1

below about 1 in 250 for the other three models. Most of these departures are attributable to three athletes: one who had high ON-he scores at all three visits, one who had extremely high OFF-hre and OFF-hr scores at two of his three visits, and one with hereditary spherocytosis who had extreme ON-he scores at all three visits and extreme OFF-hre and OFF-hr scores at one visit. Without these three athletes, the adjusted and theoretical false-positive rates were in close agreement.

The relative ability of the models to detect rHuEPO use during different phases of rHuEPO administration is illustrated in Figure 2 (for false-positive rates of 1 in 100 for the ON models, and 1 in 1,000 for the OFF models). Sensitivities overall were clearly lower for all models in the Oslo trial than in the Canberra trial. In both trials the old ON model had generally higher sensitivity during the period of high-dose administration, but the new models had higher rates during the low-dose period. The ON-hes model had a little more sensitivity than the ON-he model for both males and females.

The new OFF models were clearly superior to the old OFF model, both in the Canberra trial and also in the Oslo trial (where the previous model had zero sensitivity in all tests for a false-positive rate of 1 in 1,000). There was little difference between the two new models in the Canberra trial, but in the Oslo trial the OFF-hre model was somewhat more sensitive overall than the OFF-hr model.

Figure 3 shows the effect of choice of false-positive rate on the ability of the models to detect rHuEPO administration using cut-offs appropriate for worst-case athletes tested at sea level. Two sets of points are plotted in each panel, one set for worst-case athletes tested at sea level and one for typical athletes tested at sea level. The sensitivities for worst-case athletes are higher than those for typical athletes since, by definition, model scores for worst-case athletes are, on average, higher than those for typical athletes and hence users amongst worst-case athletes are more likely to produce a model score that exceeds a given cut-off. Points for a given athlete group form part of a sigmoidal curve with asymptotes at 0% and 100% sensitivity. For typical athletes of either sex tested at sea level (the open circles in Figure 3) and receiving a low dose of rHuEPO (as in the Oslo study) detection with a sensitivity of about 50% would be associated with a false-positive rate of about 1 in 20 for the ON-hes model and about 1 in 10 for the ON-he model. For a higher rHuEPO dosage (as used in the Canberra study), the false positive rate associated with a sensitivity of about 50% with either ON model would be about 1 in 200.

Sensitivities for the OFF models were higher. The OFF-hre model achieved detection rates of about 50% with a false-positive rate of 1 in 100 (males) and 1 in 20 (females) after low-dose injections had ceased. After the cessation of a higher dose of rHuEPO (as in the Canberra trial), the same sensitivities were achieved with false-positive rates of 1 in 10,000 (males) and 1 in 1,000 (females). For OFF-hr, the same sensitivities were obtained with a false-positive rate that was approximately double those reported for the OFF-hre model.

Sensitivities for worst-case athletes tested at sea level (the filled circles in Figure 3) were substantially higher than for typical athletes. The reason for this is that model scores for typical athletes tend to be smaller and therefore have to be increased by a larger amount (by taking rHuEPO) in order to exceed the cut-offs appropriate for worst-case athletes. Rates of detection of 50% were achieved with the ON models with false-positive

rates of 1 in 20 to 1 in 100 using low doses, and around 1 in 1,000 when using higher doses of rHuEPO. With the OFF models, the corresponding false-positive rates were 1 in 200 to better than 1 in 10,000. Cut-off scores corresponding to various false-positive rates for typical and worst-case endurance athletes with each of the models are shown in Table 2. Cut-offs for worst-case athletes tested at altitude are also shown in Table 2.

Discussion

The principal outcome from our research is the development of second-generation blood tests that possess a range of favorable characteristics that could leverage the efficacy of current antidoping strategies.

Two important characteristics are enhanced sensitivity within several days of an injection with moderate to low doses of rHuEPO, and enhanced sensitivity for up to three weeks after injections cease.

The former confers substantial benefits for the screening of urine samples, the latter for identifying athletes who recently ceased using rHuEPO and therefore warrant follow-up testing. The capacity of our new models to detect athletes using low doses of rHuEPO provides the opportunity to detect athletes even when they are using *maintenance* doses of rHuEPO.

Using blood models to screen for urinalysis

At present some authorities elect to analyze urine samples from athletes whose blood sample demonstrates unusually high Hb levels, or perhaps a combination of high Hb and elevated reticulocytes – based on the assumption that such athletes are more likely to have used rHuEPO. These characteristics are typical of rHuEPO abuse, and can be conveniently derived during field-based analysis of blood collected for *health checks* on the day of competition. Whilst such a strategy would be sensitive if the athlete was using rHuEPO to *elevate* their red cell mass, once their target red cell mass had been achieved they would be obliged to reduce their dosage or risk either cardiovascular complications associated with thickened blood and/or the health-based exclusion criteria used by some sporting federations. Maintaining a supra-normal red cell mass requires only that those red cells destroyed each day are replaced, and this is difficult to distinguish from the normal or *basal* level of red cell production encountered in most athletes. This implies that a reticulocyte-sensitive technique would fail to identify an athlete using *maintenance* doses of rHuEPO because reticulocytes are not elevated during these doses.[12]

The data from the Oslo trial portray the hematologic profile encountered during this *maintenance* phase. As depicted in Figure 2, the ability of our previous ON-model to detect users once the dosage had been reduced to a maintenance dosage (days 22-47) was poor. However our new ON-hes and ON-he models were capable of detecting approximately 40% of these subjects despite their reticulocyte count returning to baseline levels during the maintenance phase. The robustness of this superior sensitivity was demonstrated during the Canberra trial (Figure 2) in which the sensitivity of the new models was more or less sustained despite a substantial decline in the sensitivity of the previous ON-model during the maintenance phase.

Based on the additional information that serum parameters can provide, the federation may seek to supplement the preliminary knowledge derived from the Hb and reticulocyte measures in order to identify those samples with an increased likelihood of containing recombinant erythropoietin (and therefore optimize their expenditure on urine testing). An aliquot of serum could be measured for EPO and sTfr, both of which are transiently increased immediately following an injection of rHuEPO. This would enable the federation to calculate an ON-hes model score, which is elevated only for the 3-4 days following a rHuEPO injection (*unpublished observations*) and therefore complements the window of sensitivity of the urine test (both tests are sensitive to the presence of rHuEPO in the system). Establishing the position of the ON-hes model score on Figure 1 would inform the laboratory of the likelihood that the athlete had taken rHuEPO in recent days. Such an approach can substantially reduce the expense of testing a large group of athletes for rHuEPO use, as illustrated in the following cost-benefit analysis.

Several assumptions have been made for the purpose of this example: estimated costs of blood ($60) and urine ($400) analyses are exclusive of sample collection costs; within a cohort of 1,000 athletes, 20 of these are using rHuEPO; only 50% of these users will have sufficient erythropoietin in the urine (measured via a preliminary urine assay) to enable successful detection with the urine test, but providing there is sufficient erythropoietin in the urine, the electrophoretic technique for rHuEPO is 100% sensitive; and finally the blood model has a sensitivity of 80% and a false positive rate of 1 in 10. Considering the scenario in which the federation relies only upon urine analysis, approximately 500 samples would be analyzed (since only about half of all urine samples would have sufficient erythropoietin to warrant the analysis), and 10 of the 20 athletes would be caught – at a cost of $200,000. Alternatively, if a blood model with 80% sensitivity and 1 in 10 false positive rate was used as a preliminary screen on the 1,000 samples (cost $60,000), it would flag 114

athletes as unusual, comprising 98 (10% of the 980) non-users and 16 (80% of the 20) users. Because only 57 of the 114 samples would have sufficient urinary erythropoietin to warrant analyses, urine testing of these samples would cost a further $22,800, and would catch 8 of the 16 users. The total cost would be $82,800 to catch 8 users.

Selection of blood parameters for new models

The concept of using multiple-parameter models to detect rHuEPO abuse was initially explored by Australian researchers during a pilot study at the Australian Institute of Sport in 1999.[13] Subsequent funding and support were based upon these preliminary findings, and required demonstration of the reproducibility of our initial results. Since our laboratory had access to only one hematology analyzer, all subsequent work was based on this technology. This inevitably led to a significant challenge apparent today – how to interpret and apply data collected on instruments other than the machine used to derive the initial models. Mindful of this limitation, our new models were conceived using an iterative approach, which considered both the physiologic disturbances associated with rHuEPO administration, and an appreciation of the suitability/availability of various hematologic parameters for use in an antidoping setting. We sought to develop second-generation models using parameters with universal acceptance, however ultimately, our choices were governed by the sensitivity of parameters to (sub-clinical) alterations in the rate of erythropoiesis, and this mandated the inclusion of serum measures of sTfr and EPO, as well as percent reticulocytes to express immature red blood cell release.

We believe the enhanced sensitivity associated with inclusion of a third parameter (sTfr for ON-hes, EPO for OFF-hre) justifies the added expense of conducting these assays. However the cost-benefit analysis of measuring two versus three (or whole blood only versus whole blood plus serum assays for the OFF models) should ultimately be undertaken by the entity applying these models.

Regarding a parameter to reflect oxygen carrying capacity, we gave consideration to the debate surrounding the suitability of either Hct or Hb as the principal expression of this characteristic. d'Onofrio has argued persuasively that Hb is the preferred choice because it is the more precise, accurate, direct and standardized blood parameter.[14]

Hemoglobin concentration, obtained by photometric measurement, is supported by a unique reference preparation and by an established reference method accepted by the *International Council for Standardization in Hematology*. Hematocrit, on the other hand, has several undesirable characteristics that render it less than ideal to be used in an anti-doping setting. The *gold standard* for hematocrit is the volume ratio of red cells to whole blood after centrifugation, whilst the automated method relies on electrical impulses generated by the cell when it passes through impedance-based or optical-based analyzers. A false increase due to plasma trapping in the corner of red cells during centrifugation causes discrepancies between the manual and automated method. This discrepancy cannot be removed by a correction factor since the effect is non-linear.[14] Although both Hct and Hb are sensitive to plasma volume changes, only the former is sensitive to changes in mean cell volume which can occur during extended storage/transportation.

Interpretation and application of new models

The ability to tailor the specificity and sensitivity of the chosen model by selecting a cut-off threshold greatly enhances the potential application of our blood models. However, in our previous paper, we demonstrated that various fixed effects influenced the erythropoietic-sensitive parameters included in our blood models.[11] In our new models, the primary influences were the effects of altitude, ethnicity, and endurance training on Hb. After these fixed effects were accounted for, the observed rate of false-positives was comparable with the theoretical rate (the minor departures that were observed can be accounted for by the model scores from three athletes, one of whom had been diagnosed as having hereditary spherocytosis). Nevertheless, in practice it may be problematic to establish which fixed effects should be allowed for when calculating the model score for a particular athlete's blood sample – the anonymity of subjects' samples, a desirable aspect of current antidoping practice, seems to preclude the provision of adequate information to the analytical laboratory.

One approach that negates the need to access information on subjects' characteristics is to modify the cut-off score. However increasing the cut-off score (which has the same consequence as incorporating a fixed effect into the model score) reduces the ability to detect rHuEPO users (decreased sensitivity). Because the decision as to where to set the cut-off has such important implications, and is ultimately influenced by a multitude of subjective considerations, we deemed it desirable in the current paper to provide readers with sufficient comparative information to allow them to make informed decisions regarding which hematologic parameters and cut-off scores best suit their intended application.

Furthermore, the effect of altitude is unambiguous – if the blood sample was collected from the athlete whilst at altitude, a compensation should be made for this. Further research is required to establish the persistence of this hemoconcentra-

tion when the athlete is tested at sea level, and whether any allowance should be made for athletes who have recently been to altitude.

Blood as an antidoping matrix

In addition to the capacity of blood monitoring to rationalize the cost of urinalysis, and identify low-dose users, it also offers a powerful tool against rogue athletes seeking to avoid sanction. The most obvious tactic for an athlete seeking to avoid detection would be to adopt novel doping strategies – this tendency has been ratified by the appearance of Aranesp as a doping agent soon after the urine test for first-generation EPOs was introduced. The ability of Lasne and de Ceaurriz's electrophoretic technique[1] to also detect Aranesp stands as testimony to the versatility and robustness of their approach. However it is equivocal whether their test, which identifies exogenous rHuEPO based on isoelectric profiles, will be effective against next-generation EPOs derived from human cell lines (that may be very similar to endogenous EPO). In contrast, any doping agent that stimulates red cell production will be susceptible to a blood model sensitive to erythropoietic parameters.

However this deterrent umbrella is opened further by the incorporation of an OFF model. Were an athlete to cease injecting rHuEPO, either because they travelled to compete across national borders and feared criminal prosecution if caught in possession of an illicit substance, or could not afford the expense of year-round rHuEPO injections, or even because they ceased using rHuEPO in the lead up to a major event for fear of unannounced testing, their prior usage would be betrayed by an elevated OFF model score. Provided that the athlete was required to submit a blood sample at the time of competition (either for a *health check* or conceivably as part of a routine doping control if they obtained a podium finish), the routine Hb and reticulocyte parameters could be used to calculate an OFF-hr score. In response to the unusual score, the federation might seek a genetic/environmental explanation. In the absence of mitigating circumstances explaining why the athlete possessed a blood profile typical of recent/discontinued blood doping, the athlete should at least be requested to participate in a medical evaluation to ascertain the basis of their abnormal score. This follow-up may detect an undiagnosed medical condition (although it should be noted that the hematologic milieu of increased Hb together with abnormally low reticulocyte and EPO levels has not been ascribed to any known pathological abnormality in the literature), or provide the federation with additional data to help recognize the probable reason for the elevated OFF model score.

The capacity of the new OFF models to highlight an athlete who has used rHuEPO is substantial. For example, using our OFF-hre model, only 1 athlete in every 1,000 is likely to exceed the (theoretical) 1 in 1,000 cut-off score; however the associated sensitivity graph demonstrates that at various times 20-80% of subjects who cease taking rHuEPO have a model score above this threshold. This ability to identify dopers should be weighed against the likelihood of encountering an athlete who possesses this model score as a consequence of natural biological variability. The subjective decision as to what level of false-positive risk a participant is willing to accept in order to have recourse to a sensitive tool capable of identifying their competitors who are cheating, arguably belongs to the athletes themselves. Sports governing bodies should heed the wishes of their constituent athletes, and perhaps the debate on whether or not to implement blood monitoring could be the responsibility of comprehensively-briefed athlete commissions.

References

1. Lasne F, de Ceaurriz J. Recombinant erythropoietin in urine. Nature 2000;405:635.
2. Parisotto R, Wu M, Ashenden MJ, Emslie KR, Gore CJ, Howe C, et al. Detection of recombinant human erythropoietin abuse in athletes utilizing markers of altered erythropoiesis. Haematologica 2001;86:128-37.
3. Genomics-based Drug Data Report and Regenerative Therapy 2001. www.prous.com/journals/gddr/sample/html. 1:5-19.
4. Egrie JC, Brown JK. Development and characterisation of novel erythropoiesis stimulating protein (NESP). Br J Cancer 2001;84:3-10.
5. Birkeland KI, Stray-Gundersen J, Hemmersbach P, Hallen J, Haug E, Bahr R. Effect of rhEPO administration on serum levels of sTfR and cycling performance. Med Sci Sports Exerc 2000;32:1238-43.
6. Audran M, Gareau R, Matecki S, Durand F, Chenard C, Sicart MT, et al. Effects of erythropoietin administration in training athletes and possible indirect detection in doping control. Med Sci Sports Exerc 1999;31:639-45.
7. Bressolle F, Audran M, Gareau R, Pham TN, Gomeni R. Population pharmacodynamics for monitoring epoetin in athletes. Clin Drug Invest 1997;14:233-42.
8. Gareau R, Audran M, Baynes RD, Flowers CH, Duvallet A, Senecal L, et al. Erythropoietin abuse in athletes. Nature 1996;380:113.
9. Cowart VS. Erythropoietin: a dangerous new form of blood doping? The Physician and Sports Medicine 1989;17:115-8.
10. Woodland L. Lethal Injection. Bicycling 1991;32:80-81.
11. Sharpe K, Hopkins W, Emslie KR, Howe C, Trout GJ, Kazlauskas R, et al. Development of reference ranges in elite athletes for markers of altered erythropoiesis. Haematologica 2002; 87:1248-57.
12. Russell G, Gore CJ, Ashenden MJ, Parisotto R, Hahn AG. Effects of prolonged low doses of recombinant human erythropoietin during submaximal and maximal exercise. Eur J Appl Physiol 2002;86:442-9.
13. Parisotto R, Gore CJ, Emslie KR, Ashenden MJ, Brugnara C, Howe C, et al. A novel method utilizing markers of altered erythropoiesis for the detection of recombinant human erythropoietin abuse in athletes. Haematologica 2000;85:564-72.
14. d'Onofrio G. Hematology parameters: usefulness and limitations in monitoring red cell production. Bloodline Rev 2001;1:5-6.

Pre-publication Report & Outcomes of Peer Review

Contributions and Acknowledgments

CJG: conception, design, analysis and interpretation, drafting and review, final approval; RP conception, design, analysis and interpretation, drafting and review, final approval; MJA: conception, design, analysis and interpretation, drafting and review, final approval; AGH: conception, design, analysis and interpretation, review, final approval; JS-G: conception, design, analysis and interpretation, review, final approval; KRE: design, analysis and interpretation, review, final approval; GJT: design, analysis and interpretation, review, final approval; CH: design, analysis and interpretation, review, final approval; RK: design, review, final approval; KS: design, analysis and interpretation, drafting and review, final approval; WH: analysis and interpretation, drafting and review, final approval. Responsibility for Table 1: CJG; Figures 1, 2 and 3: CJG, MJA, KS, WH; for Table 2: KS, WH.

The authors acknowledge the Australian Government and the International Olympic Committee for providing funds for this research. Nicole Horvath, Simone Ransley, Sally Wright, Graeme Allbon, Julianne Brettargh, and Kelly Linaker for technical support. Robyn Power, Gabrielle Russell, and Terry Gathercole (dec) for management and administrative support. Gaenor Edwards, Dorothy Manocha, Katina Urlich, Jenny Lenihan for medical support. We wish to thank Janssen-Cilag for donation of the Eprex for this study. We also wish to thank Dade Behring for donating the nephelometer and the consumables for the sTfr measurements, and Bayer Diagnostics for their technical and engineering backing with the ADVIA analyzer. Special recognition is due to the subjects who volunteered for this study.

Funding

This study was jointly funded by the Australian Federal Government and the International Olympic Committee.

Disclosures

Conflict of interest: none.

Redundant publications: no substantial overlapping with previous papers.

Manuscript processing

This manuscript was peer-reviewed by two external referees and by Dr. Carlo Brugnara, Deputy Editor. The final decision to accept this paper for publication was taken jointly by Dr. Brugnara and the Editors. Manuscript received October 29, 2002; accepted January 25, 2003.

In the following paragraphs, Dr. Brugnara summarizes the peer-review process and its outcomes.

What is known in this field

With the Sydney 2000 Olympic Games, a method utilizing a combination of biochemical and hematological parameters indicative of increased erythropoietic rate was implemented for detection of surreptitious use of recombinant human erythropoietin (r-HuEPO).

What this study adds

The authors provide evidence for the ability of a new simplified model (ON model) based on increased hemoglobin, serum erythropoietin level ± serum transferrin receptor levels to detect recent r-HuEPO abuse. Models to detect a depressed erythropoietic activity following r-HuEPo abuse (OFF model) are also presented: these models rely on increased hemoglobin and decreased % of reticulocytes ± decreased serum erythropoietin.

Caveats

More work needs to be done to establish the effect of altitude training, and to validate these models across various instrument and reagents platforms. These models need to be incorporated into a systematic approach which follows selected biochemical and hematological parameters over time in competitive athletes (the Hematologic Passport).