Expert opinion concerning Mrs Pechstein / ISU

Summary

It is shown that the prosecution has evaluated the evidence by assuming 'guilt', and in doing so, the apparent value of the evidence is affected. In particular, the evidence has *not* been evaluated in the potentially correct way, namely according to the principles of the biological passport, as these have been invariably spread through various channels (scientific literature, presentations, web sites). To a basic scientist, this finding is highly surprising. Further comments are provided to reveal questionable science and logical flaws on the side of the prosecution.

N.M. Faber, Ph.D.

Beek-Ubbergen, 13 October 2009

Table of contents

1. Introduction

Throughout this report, claims are substantiated by quoting relevant scientific articles, in particular from anti-doping researchers [1-4], as well as easily accessible sources such as Wikipedia. The latter, 'less respectable' sources have been selected to make certain comments sufficiently understandable for non-scientists.

It is stressed that much of the scientific work leading to the biological passport has been sponsored by the World Anti-Doping Agency (WADA) (e.g. [1] and [3]) and the International Cycling Union (UCI) (e.g. [2]). This funding can be seen as compromising impartiality.

To avoid cluttering up the main text, a great deal of the supporting material is deferred to Appendices.

2. Comments on arguments advanced by the expert on methodology (Sottas)

These comments are divided in two categories. First (in 2.1), it is shown that the current treatment of evidence does not follow from the concepts and principles underlying the biological passport, as these have been firmly consolidated in (most notably) the scientific literature [1-4]. Second (in 2.2), some additional comments are made of a more general nature, to provide a background for the specific comments made in 2.1.

2.1. The biological passport

It is stressed that completeness is not pursued in the current treatment of the biological passport. Here, the purpose is to detail the contradiction between established theory (in 2.1.1) and the current treatment of evidence (in 2.1.2 and 2.1.3).

Throughout, it should be borne in mind that the biological passport is based on *indirect* evidence.

2.1.1. Established theory: to bring reliable evidence in front of a disciplinary panel

In what is perhaps the most detailed exposition of the principles and concepts underlying the biological passport [3], Dr. Sottas claims to pursue the following:

> "The aim is to allow anti-doping organizations to bring reliable evidence of blood manipulation in front of a disciplinary panel."

The rationale is quite forceful and well consolidated in the scientific literature [1-4]. Two stages can be distinguished:

- During the first stage, data are collected for an athlete over a period when (s)he can be assumed to be 'clean'. These data serve to calculate a historical baseline that is tailored to the specific athlete. In other words, personal limits replace population limits that are used in the traditional approaches that are to be replaced by the biological passport.

- During the second stage, new data are then compared with the personal limits and deviations are evaluated according to a statistical decision rule.

Plainly put for a lay person: with an individual historical baseline, the numbers obtained in the second stage *speak for themselves*.

It is reiterated that *population limits* are replaced by *individual limits*. This requires that a period be identified where the athlete can be assumed to be 'clean'. Verbal explanations and graphical illustrations of this basic condition are given in Appendix 1.

2.1.2. Current incongruous practice: the evidence is evaluated by assuming 'guilt'

The established theory should be contrasted with the current treatment of the evidence, where *all data are compared with a population limit*. This is immediate from the graphic given in Appendix 1.

It stands to reason that by *not* identifying a period for which the athlete can be assumed to be 'clean', one effectively accepts the mutually exclusive hypothesis of 'guilt'.

Proper evaluation of evidence respects the legal principle *in dubio pro reo*. By intentionally or unintentionally assuming 'guilt', Dr. Sottas has violated this principle. Dr. Sottas should provide a valid explanation for not properly evaluating the evidence.

2.1.3. Current incongruous practice: the need for additional tests

It is reiterated that, with the development of the biological passport, Dr. Sottas claims to pursue the following [3]:

> "The aim is to allow anti-doping organizations to bring reliable evidence of blood manipulation in front of a disciplinary panel."

How to best illustrate that this ambitious aim has not been achieved in the current case? Or, plainly put: that the (current) numbers *do not speak for themselves*?

The rather sad answer to this question is provided by Dr. Sottas himself in his Supplement dated 28 August 2009:

> "However, I would suggest to envisage one of these forms of blood doping only if there is strong evidence that the athlete does not present any clinically silent blood disorder. Therefore, I would strongly suggest to perform additional tests to exclude that possibility."

It is reiterated that correct application of the principles and concepts underlying the biological passport would have ruled out "that possibility" from the very beginning.

2.2. Further comments

2.2.1. No proper underpinning of the acceptable probability of a false-positive conclusion

In scientific articles on the biological passport, one invariably encounters 1 in 1,000 as the acceptable probability of a false-positive conclusion [1-3]. Interestingly, this value is *a factor of 10 higher* than the one considered by Sonksen [5], see:

> "The sports authorities have never published what they consider an 'acceptable risk' (for a false +ve) but a workshop in our GH-2000 project that included a senior IOC lawyer settled on a risk of 1:10,000 as being 'acceptable' and this figure has subsequently been used in many discussions and publications."

Reliable estimates about the number of false-positives actually achieved in practice do not exist, for the obvious reason that false-positives are not identified. However, a reasonable indication may have been provided by anti-doping researchers Van Eenoo and Delbeke, who note in [6] that the number of contested positive A-sample results ranges between 0.5% and 1% (long term averages: 1996-2007). Considering that one obtains a positive result for approximately 2% of the A-samples, one arrives by simple multiplication at the probability of an A-sample being contested ranging between 1 in 10,000 and 2 in 10,000. This range should provide a reasonable upper bound for the number of false-positives actually achieved: with more (true) false-positives, more people would likely complain.

It follows that even the probability advocated by Sonksen is relatively high. If the target value 1 in 10,000 were actually achieved in practice, a substantial part of the contested A-samples would be false-positives.

It also follows that the probability chosen in the context of the biological passport could easily lead to a tenfold increase of the number of false-positives, over the ones currently achieved. This particular choice therefore seems to be *irresponsible*.

2.2.2. No proper calculation of the probability of a false-positive conclusion for the AO

In the light of the rather idiosyncratic value for the acceptable risk of a false-positive conclusion (1 in 1,000), it is perfectly understandable that, in their Report dated 31 July 2009, Profs. Kiesewetter and Röcker prompt Dr. Sottas to provide a calculation. It is well known that the proper calculation requires two test characteristics (specificity and sensitivity), as well as a population characteristic (prevalence of doping). A worked example is provided in Appendix 3.

Surprisingly, Dr. Sottas has not made any effort to perform the necessary calculation. Rather, he prefers to refer in his Supplement dated 28 August 2009 to a search in Google Scholar (see 4.1). This

is all the more awkward since Dr. Sottas states, in the same document, that prevalence is low (see 2.1.2), but how low is it?

It is important to note that Dr. Sottas, in his scientific work, displays a strong interest in prevalence [3]:

> "We have recently stressed the importance of statistical measures such as the prevalence and predictive positive values (PPV) in the interpretation of anti-doping tests (Sottas *et al*., 2006, 2007a,b). The PPV is considered as the gold standard for medical diagnostic testing because this statistics provides the probability that a positive test actually reflects the underlying condition that is being tested for (Altman & Bland, 1994)."

It follows that there can be no excuse for not providing this essential calculation.

2.2.3. Resorting to the argument from ignorance

As pointed out by Profs. Kiesewetter and Röcker in their Report dated 31 July 2009,

> "Dr. Sottas *cannot explain* the longitudinal profile of the haematological data in an elite athlete and suspect's a manipulation of results via haemodilution." (The italics are by the current author.)

It stands to reason that this logical flaw (see Appendix 4) should not be rewarded as a support of the conclusion 'guilt' – to any degree.

2.2.4. A school example of confirmation bias

From all the variables that have been measured for the AO, only a single one has been selected namely the one that could point in the direction of blood manipulation. This is a typical example of confirmation bias. It is stressed that forensic scientists are trained to avoid this type of bias.

Prof. Jelkmann, for example, has already strongly argued against the unscientific nature of this biased selection, and rightly so. For further discussion of this type of bias, see Appendix 5.

3. Further comments on arguments advanced by the expert on subject matter (d'Onofrio)

The arguments of Prof. d'Onofrio have been addressed in detail by Profs. Kiesewetter, Röcker, Heimpel, Jelkmann and Gassmann. In short, the conclusion of Prof. d'Onofrio ('blood manipulation') is not supported by the data. Illustrative is the thoughtful wording of Prof. Gassman:

> "I consider my conclusion as beyond reasonable doubt, unless assumption A or B (see point 25) turn out to be wrong."

Prof. d'Onofrio should give proper attention to these assumptions, namely (A) correct data and (B) undiagnosed (genetic) blood disease.

Moreover, even Dr. Sottas recognizes the need for *additional tests*, see 2.1.3.

Therefore, only two short *further* comments are deemed necessary here. The first one formalizes an earlier critique, while the second one follows from the detailed treatment of the biological passport in Section 2.1 and Appendices 1 and 2.

3.1. Resorting to the argument from ignorance

As pointed out by Profs. Kiesewetter and Röcker in their Report dated 31 July 2009,

"Prof. Giuseppe d'Onofrio *cannot explain* the changes in the Reticulocytes (%) and haemoglobin via a physiological mechanism. The only explanation given is that it may be the result of a light membrane defect in the erythrocytes." (The italics are by the current author.)

It stands to reason that this logical flaw (see Appendix 4) should not be rewarded as a support of the conclusion 'guilt' – to any degree.

3.2. Misunderstanding of the principles behind the biological passport

In his addition to the Expert Opinion dated 25 May 2009, Prof. d'Onofrio states:

"The athlete's biological passport, which is the present and the future of the fight against blood doping, is founded on scientific grounds (such as interindividual and intraindividual variances) which have a very sound and solid scientific basis."

From statements like this, one becomes painfully aware that Prof. d'Onofrio blindly trusts that the (only) material evidence has been properly evaluated, which is not the case.

4. Comment on Statement of complaint

I will limit myself to a single but crucial section.

Section 4.7 starts with:

"There is a solid statistical basis for concluding that the results found for the Alleged Offender establish the Use of a Prohibited Method."

From statements like this, one becomes painfully aware that the ISU blindly trusts that the (only) material evidence has been properly evaluated, which is not the case. Without an estimated probability of a false-positive conclusion, constructs like "solid statistical basis" have no meaning at all.

Section 4.7 continues with:

"The chance for a single value above the upper cut off level of 2.4% is approximately 1%. The chance for a value of percent reticulocytes above 3% is approximately 1‰."

Here, the following quote from [3] applies:

> "2.1 First paradigm: detection of an abnormal sample
>
> An athlete may be excluded from a competition if the analysis of his/her blood sample collected just prior the competition reveals an abnormal value for a blood parameter. In this paradigm, the decision rule is based on an *arbitrary threshold* of the specificity of the blood parameter, and *not on a true evidence of blood manipulation*." (The italics are by the current author.)

It is one of many contradictions in the treatment of the current evidence.

5. Concluding remarks

Numerous conclusions come to mind, but I will limit myself to the following:

- It has become painfully clear that Prof. d'Onofrio as well as the ISU blindly trust that the (only) material evidence has been properly evaluated, which is not the case. By assuming 'guilt', the apparent value of the evidence is affected.
- The biological passport is presented as a *forensic* approach, see e.g. [3,4]. However, this claim does not hold true. By contrast, anti-doping researchers should seriously consider adopting a system of widely accepted forensic practices, as has recently been argued [7]. In fact, Sally Clark would be convicted again if the (only) indirect evidence were evaluated in the manner displayed here. In more ancient times, Josyne van Beethoven was the victim of highly convincing indirect evidence. Although nobody will deny that considerable progress in science has since then been achieved, one should not overlook the same psychological mechanisms repeatedly being at work, exemplified by argumentation from ignorance, confirmation bias, etc.

6. Implications

The availability of a single piece of material evidence, which, owing to the lack of an individual historical baseline, by itself is *grossly incomplete*, implies that:

- A reasonable standard of proof has not been met.
- The burden of proof is (still) upon the ISU.

7. Recommendations

Numerous recommendations come to mind, but I will limit myself to the following:

- From various points of view, additional tests could be recommended. However, one should ask oneself whether it is desirable to continue violating basic human rights to 'reward' flawed science.
- It is tempting to speculate on reasons for treating the evidence in a way that clearly contradicts previous scientific descriptions [1-4]. Dr. Sottas should be prompted to provide a valid explanation.

Appendix 1: The biological passport, as it *was* firmly consolidated

Here follow a number of quotes that clearly demonstrate that proper application of the concepts behind the biological passport requires identification of a period for which the athlete is assumed to be 'clean'. All italics in these quotes are by the current author.

For a scientist with a proper background, from [1]:

> "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*. 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*."
>
> (…)
>
> ***"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*."

For a scientist with a proper background, from [2]:

> "Along with this, the idea of a hematological passport was also suggested. Sportsmen with *significant differences between new test results and an individual historical baseline* could be excluded from competition."

For a lay person, from [8]:

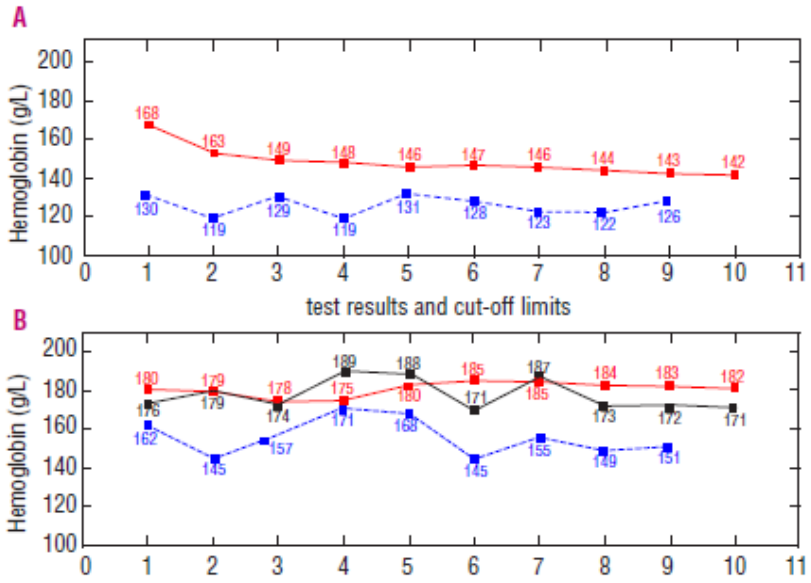
> **"What is a haematological profile?**
>
> This is the newest element on the UCI's anti-doping program.
>
> The main difference is that a series of tests from each rider will be organised into a profile which enables *individual limits for each rider* to be established. The current anti-doping approach compares each blood sample to a population limit. The population limit currently determines the "no start" rule. *With the new approach, each sample will be compared with the rider's own individual "normal" haematological levels. Any significant variations can then be assessed for possible blood manipulation*."

For a lay person, from [9]:

> "A biological passport is an individual, electronic record for each professional racing cyclist, in which the blood profiles and results of all doping tests over a period of time are collated. *Doping violations can be detected by noting variances from an athlete's established levels* outside permissible limits, rather than testing for and identifying illegal substances."

Since a picture says more than a thousand words, two graphical displays follow.

The following graph is taken from [2]:



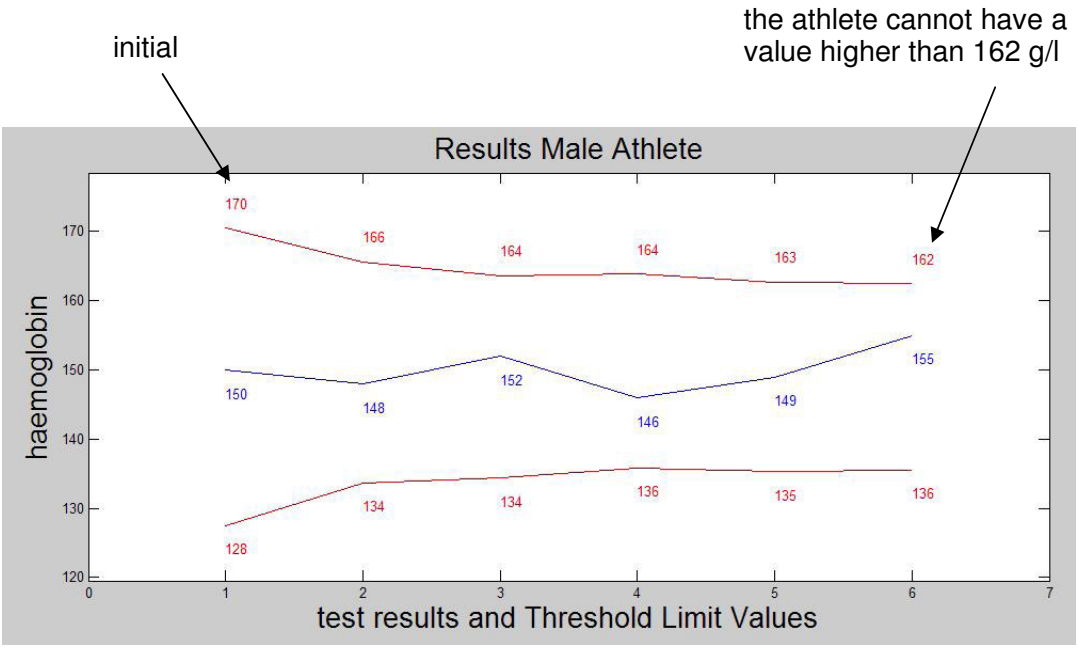
Explanation upper panel (A), quoted from [2]

"Red line represents the limits found for a specificity of 99.9% for the female subject. The first value at 168 g/L represents a population threshold: 1 in 1,000 28 year old female Caucasian endurance athletes living at low altitude should in average present a value higher than 168. The cut-off limit changes as soon as individual Hb values (blue data) are taken into account. *In case of a very high number of individual test results, the last value, here 142, represents an individual limit independent of any population factor.*" (Again, the italics are by the current author.)

Explanation lower panel (B), quoted from [2]

"Red line represents the limits for the male subject. Information about his location was not available. Black line represents limits, however, including the information that all tests were conducted at low altitude, except for the 4th, 5th and 7th tests which were conducted above 2,000 meters. Interestingly, the final cut-off limit was lower when an altitude model was explicitly considered (171 *vs* 182) i.e. a higher sensitivity to blood doping. Likewise, thresholds were higher when the athlete is tested in high altitude (e.g. 189 *vs* 175 for the 4th test) i.e. a lower probability of a false-positive."

The following graph is taken from [10].

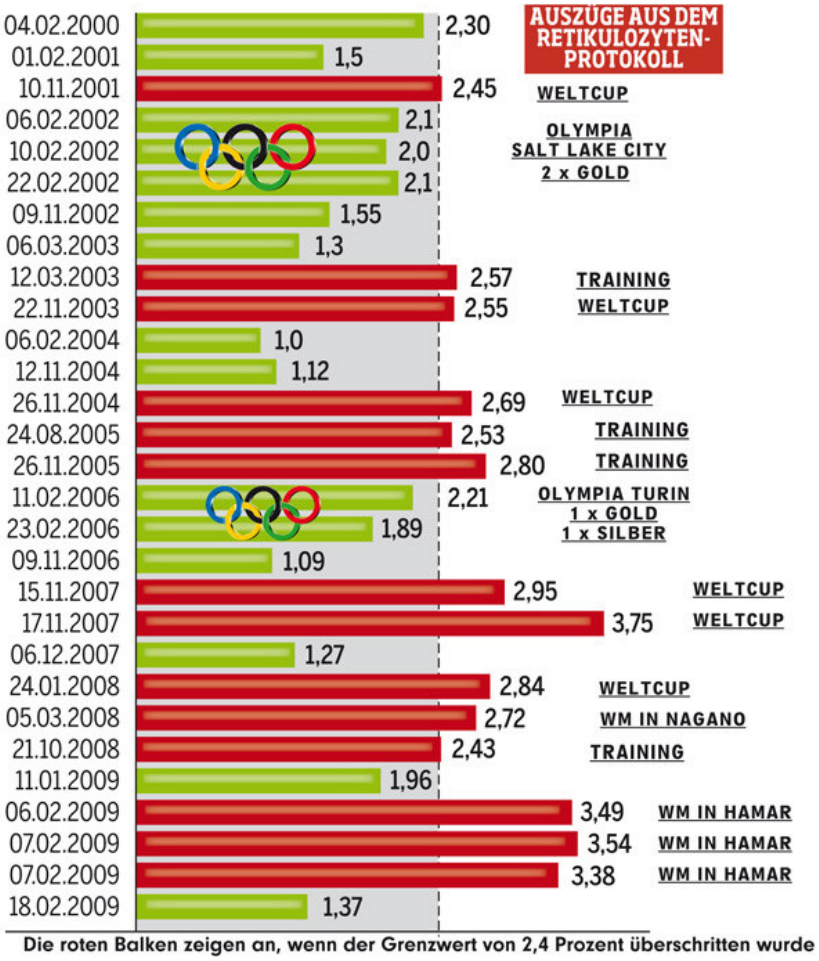


Explanation

Most certainly, it is the original intention of the biological passport to compare new values with the individual limit 162 ("the athlete cannot have a value higher than…"), not the population limit 170, which only serves to initiate the passport.

Appendix 2: The biological passport, as it *is* currently deployed

The following graph displays the reticulocyte values for the AO [11]:



N.B. For making the main point below, it is not necessary to consider that some values may have changed since this graph was published, or may have been discarded altogether (period until 20-10-2007). Also, the type of variable ('RET%') is immaterial for proper discussion of these methodological issues. It's all about numbers.

Explanation

Clearly, all values are compared with a population limit – the dashed vertical line ('Grenzwert von 2,4 Prozent') – not a personalized value.

The contrast is striking and not without consequences since *not* analyzing the data as in Appendix 1 affects the apparent value of the evidence.

Further comment

As an aside, it is noted that this population limit is probably (see slide 16 of [12]) taken from a study of Zorzoli ('Blood monitoring in anti-doping setting') [13] that ends with:

> "To conclude, we believe that abnormal blood profile cannot be considered as a proof of doping alone, because the standards for analysis are not well defined as for anti-doping purposes and because variability is very important, due to intra- and inter-individual variations and intra- and inter-technology differences, which can be very high in case of reticulocytes measurement. Nevertheless, these results can be considered as a suspicion of doping, and be used as complementary evidence during disciplinary proceedings."

For a scientist it is frustrating to see that anti-doping researchers cannot even seem to agree about what valid conclusions can be drawn from data.

Appendix 3: The proper calculation of the probability of a false-positive conclusion

The following worked example, which is particularly suited to emphasize the importance of prevalence, is taken from [14]. Please note the conclusion for this particular example that, given the positive result, "it is in fact *more* likely that the employee is *not* a drug user".

"Example 1: Drug testing

An example of the use of Bayes' theorem is the evaluation of drug test results. Suppose a certain drug test is 99% sensitive and 99% specific, that is, the test will correctly identify a drug user as testing positive 99% of the time, and will correctly identify a non-user as testing negative 99% of the time. This would seem to be a relatively accurate test, but Bayes' theorem can be used to demonstrate the relatively high likelihood of miss-classifying non-users as users. Let's assume a corporation decides to test its employees for drug use, and that only 0.5% of the employees actually use the drug. What is the probability that, given a positive drug test, an employee is actually a drug user? Let "*D*" stand for being a drug user and "*N*" indicate being a non-user. Let "+" be the event of a positive drug test. We need to know the following:

- $P(D)$, or the probability that the employee is a drug user, regardless of any other information. This is 0.005, since 0.5% of the employees are drug users. This is the *prior probability* of *D*.
- $P(N)$, or the probability that the employee is not a drug user. This is $1 - P(D)$, or 0.995.
- $P(+|D)$, or the probability that the test is positive, given that the employee is a drug user. This is 0.99, since the test is 99% accurate.
- $P(+|N)$, or the probability that the test is positive, given that the employee is not a drug user. This is 0.01, since the test will produce a false positive for 1% of non-users.
- $P(+)$, or the probability of a positive test event, regardless of other information. This is 0.0149 or 1.49%, which is found by adding the probability that a true positive result will appear (= 99% x 0.5% = 0.495%) plus the probability that a false positive will appear (= 1% x 99.5% = 0.995%). This is the prior probability of +.

Given this information, we can compute the posterior probability $P(D|+)$ of an employee who tested positive actually being a drug user:

$$\begin{aligned} P(D \mid +) &= \frac{P(+ \mid D)P(D)}{P(+)} \\ &= \frac{P(+ \mid D)P(D)}{P(+ \mid D)P(D) + P(+ \mid N)P(N)} \\ &= \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.01 \times 0.995} \\ &= 0.3322 \end{aligned}$$

Despite the specificity and sensitivity of the test, the low base-rate of use renders the accuracy of the test low: the probability that an employee who tests positive actually using drugs is only about 33%, so it is in fact *more* likely that the employee is *not* a drug user. The rarer the condition for which we are testing, the greater the percentage of positive tests that will be false positives."

Appendix 4: Argument from ignorance

The following illustrative quotes are excerpted from [15].

"The *argument from ignorance*, also known as *argumentum ad ignorantiam* ("appeal to ignorance"), *argument by lack of imagination*, or *negative evidence*, is a logical fallacy in which it is claimed that a premise is true only because it has not been proven false, or is false only because it has not been proven true."

(…)

"Commonly in an *argument from personal incredulity* or *argument from ignorance*, the speaker *considers* or *asserts* that something is false, implausible, or not obvious to them *personally* and attempts to use this gap in knowledge as "evidence" in favor of an alternative view of his or her choice. Examples of these fallacies are often found in statements of opinion which begin: "It is hard to see how...," "I cannot understand how...," or "it is obvious that..." (if "obvious" is being used to introduce a conclusion rather than specific evidence in support of a particular view)."

(…)

"The two most common forms of the argument from ignorance, both fallacious, can be reduced to the following form:

- Something is currently unexplained or insufficiently understood or explained, so it is not (or must not be) true.
- Because there appears to be a lack of evidence for one hypothesis, another chosen hypothesis is therefore considered proven."

(…)

"Law

In most modern criminal legal systems there is a presumption of innocence, and it is the responsibility of the prosecution to prove (usually "beyond reasonable doubt") that a defendant has in fact committed a particular crime. It is a logical fallacy to presume that mere lack of evidence of innocence of a crime is instead evidence of guilt. Similarly, mere lack of evidence of guilt cannot be taken as evidence of innocence."

(…)

"Also, as a hypothetical example of an "argument from personal incredulity," defined above, suppose someone were to argue:

- I cannot imagine any way for Person P to have executed action X without committing a crime Y
- Therefore, Person P must be guilty of crime Y."

(…)

Appendix 5: Confirmation bias

The following quote is taken from [16]:

> "In psychology and cognitive science, confirmation bias is a tendency to search for or interpret new information in a way that confirms one's preconceptions and to irrationally avoid information and interpretations which contradict prior beliefs. Confirmation bias is a type of cognitive bias and represents an error of inductive inference, or as a form of *selection bias* toward confirmation of the hypothesis under study or disconfirmation of an alternative hypothesis." (Italics are by the current author.)

Focusing on percentage reticulocytes *only* is the result of an apparent *selection bias*. It follows that objectivity has been severely compromised from the very beginning.

As an aside, it is noted that in (standard) anti-doping work, one attempts to avoid this kind of potentially harmful biases by analyzing the A-sample in a strictly anonymous fashion. The B-sample, however, is not anonymous and usually analyzed in the same laboratory, mostly for practical reasons like logistics.

Since disagreement of A- and B-sample results is undesirable for the laboratory in question, confirmation bias is implicitly present, but perhaps tolerable when certain precautions are carefully met (e.g. different analysts performing crucial tasks).

However, the current explicit form is rather striking and certainly not tolerable. After all, what precautions could have been met?

For example, the precautions mentioned in Section 3.3 of the Statement of complaint do not play any role in the current discussion, which is about the correct interpretation of correctly obtained numbers.

References

[1] K. Sharpe, M.J. Ashenden and Y.O. Schumacher. A third generation approach to detect erythropoietin abuse in athletes. *Haematologica*, **91** (2006) 356-363.

[2] N. Robinson, P.-E. Sottas, P. Mangin and M. Saugy. Bayesian detection of abnormal haematological values to introduce a no-start rule for heterogeneous populations of athletes. *Haematologica*, **92** (2007) 1143-1144.

[3] P.-E. Sottas, N. Robinson, M. Saugy and O. Niggli. A forensic approach to the interpretation of blood doping markers. *Law, Probability and Risk*, **7** (2008) 191-210.

[4] P.-E. Sottas, C. Saudan and M. Saugy. Doping: a paradigm shift has taken place in testing. *Nature*, **455** (2008) 166.

[5] http://www.bmj.com/cgi/eletters/337/jul04_1/a584; accessed 30 September 2009.

[6] P. Van Eenoo and F.T. Delbeke. Response on ‘‘Regulations in the field of residue and doping analysis should ensure a well-defined risk of a false positive declaration’’ by N. M. Faber. *Accreditation and Quality Assurance*, **14** (2009) 219-221.

[7] K. Faber and M. Sjerps. Anti-doping researchers should conform to certain statistical standards from forensic science. *Science and Justice*, **49** (2009) 214-215.

[8] http://www.uci.ch/Modules/ENews/ENewsDetails.asp?MenuId=&id=NTQzOA; accessed 30 September 2009.

[9] http://en.wikipedia.org/wiki/Biological_passport; accessed 30 September 2009.

[10] http://www.anado.org/documents/UCI.pdf; accessed 30 September 2009.

[11] http://www.bild.de/BILD/sport/mehr-sport/2009/07/07/claudia-pechstein/verweis-statistik,property=Download.jpg; accessed 30 September 2009.

[12] www.hetinstrument.nl/archief/2008/images/introductie.pdf; downloaded 30 September 2009.

[13] http://www.dshs-koeln.de/biochemie/rubriken/07_info/UCI_Studie.pdf; downloaded 30 September 2009.

[14] http://en.wikipedia.org/wiki/Bayes'_theorem; accessed 30 September 2009.

[15] http://en.wikipedia.org/wiki/Argument_from_ignorance; accessed 30 September 2009.

[16] http://en.wikipedia.org/wiki/Confirmation_bias; accessed 21 August 2009.