Claudia Pechstein

could she be the Lucia de Berk of elite sport?
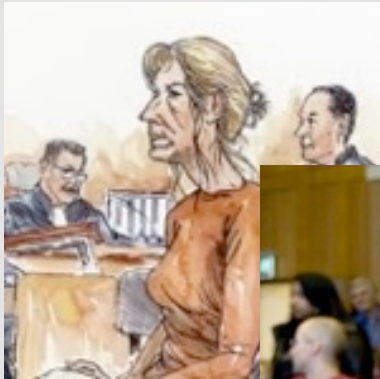

Klaas Faber

www.chemometry.com

Claudia Pechstein

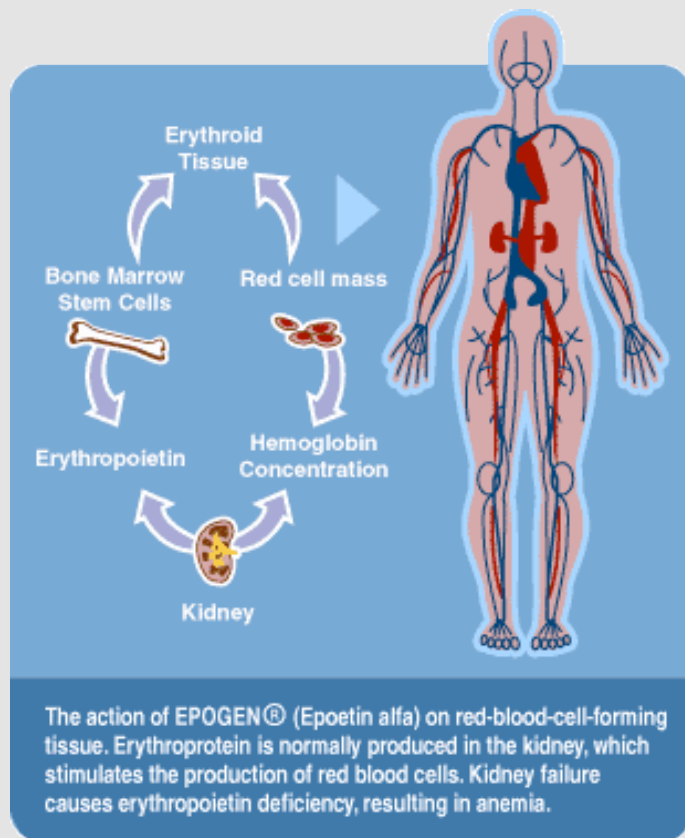**could she be the Lucia de Berk of elite sport?**

Claudia Pechstein

could she be the Lucia de Berk of elite sport?

Overview

Thursday, October 20, 2011

1. Example of blood doping: EPO

Erythroid Tissue

Bone Marrow Stem Cells

Red cell mass

Erythropoietin

Hemoglobin Concentration

Kidney

The action of EPOGEN® (Epoetin alfa) on red-blood-cell-forming tissue. Erythroprotein is normally produced in the kidney, which stimulates the production of red blood cells. Kidney failure causes erythropoietin deficiency, resulting in anemia.

Originally intended to treat patients with anemia, erythropoietin (EPO) was soon recognized as a performance enhancing means in sports.

Two stages:

1. Increased level of young red blood cells (reticulocytes);
2. Increased level of mature, oxygen-carrying red blood cells (erythrocytes).

Source: Deutsche Sporthochschule Köln

3

EPO abuse in professional cycling

N.B. Inferred from fluctuating blood values; no sanction possible.

☐ 1996: No testing ➔ more than 80%.

☐ 1997-1999: Threshold for hematocrit (% of blood volume that is occupied by red blood cells < 50) ➔ slight decrease.

☐ (…)

☐ 2003: Lack of supervision in the Vuelta (Spain) ➔ almost 100%.

☐ (…)

4

Source: Remco Geisser, 'Die Blutspur des Radsports', *NZZ*, August 5, 2007

Erythropoietin (EPO): glycoprotein
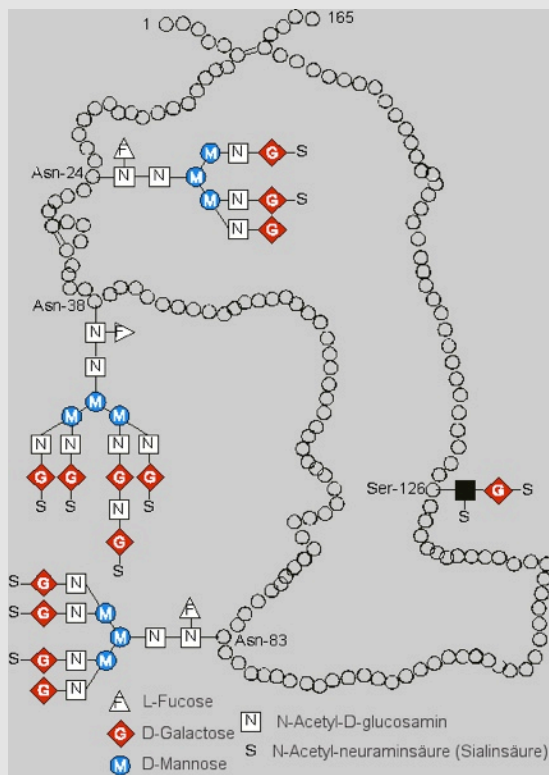
- Produced in Chinese hamster ovary cells by recombinant DNA technology.

- Various isoforms.

- Polypeptide chain molecular mass of 21,000 Da.

- Glycosylated molecular mass of 30,400 Da (28% non-protein).

Source: Deutsche Sporthochschule Köln

Thursday, October 20, 2011

Detection: gel electrophoresis

The direct detection method is based on a fingerprint of the various isoforms of exogenous EPO.

Source: Douwe de Boer, Maastricht University

Thursday, October 20, 2011
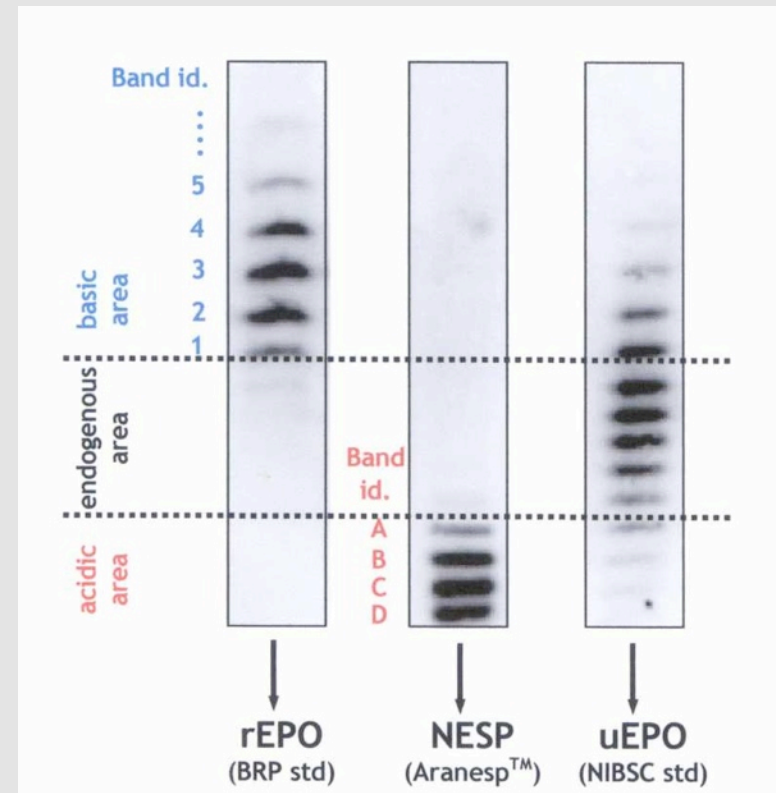
Shift focus to the effects of blood doping

☐ Conventional: direct detection of a substance in e.g. urine.

☐ Fundamental issue: limited time window and therefore limited opportunity for catching 'cheats'.

☐ Idea: expand the time window by focusing on relatively long-lived effects of doping, e.g. a rise of hemoglobin ➔ indirect evidence that requires statistics.
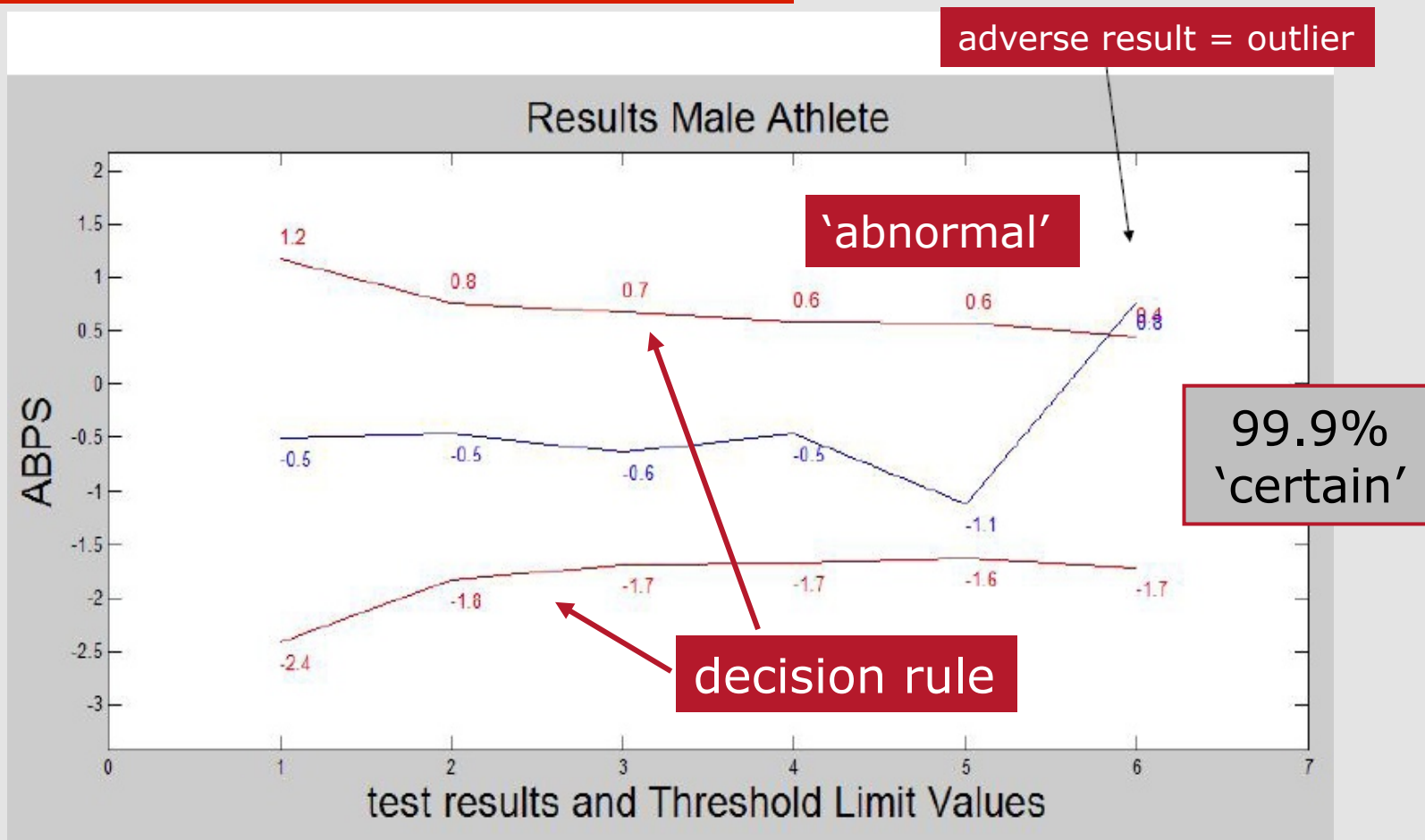

7

Prosecution as per January 1, 2009

☐ Finally, after collecting blood values for more than a decade!

☐ Comment to Article 2.2: It has always been the case that Use or Attempted Use of a Prohibited Substance or Prohibited Method may be established by any reliable means.

(...)

Use or Attempted Use may also be established by other reliable means such as admissions by the Athlete, witness statements, documentary evidence, conclusions drawn from longitudinal profiling, (...).

☐ Expectations are accordingly: more cases brought to trial!

8

Source: World Anti-Doping Code

2. The biological passport

Results Male Athlete

adverse result = outlier

'abnormal'

99.9% 'certain'

decision rule

blood value

Source: Anne Gripper (cycling union), ANADO Workshop (2008)

Thursday, October 20, 2011
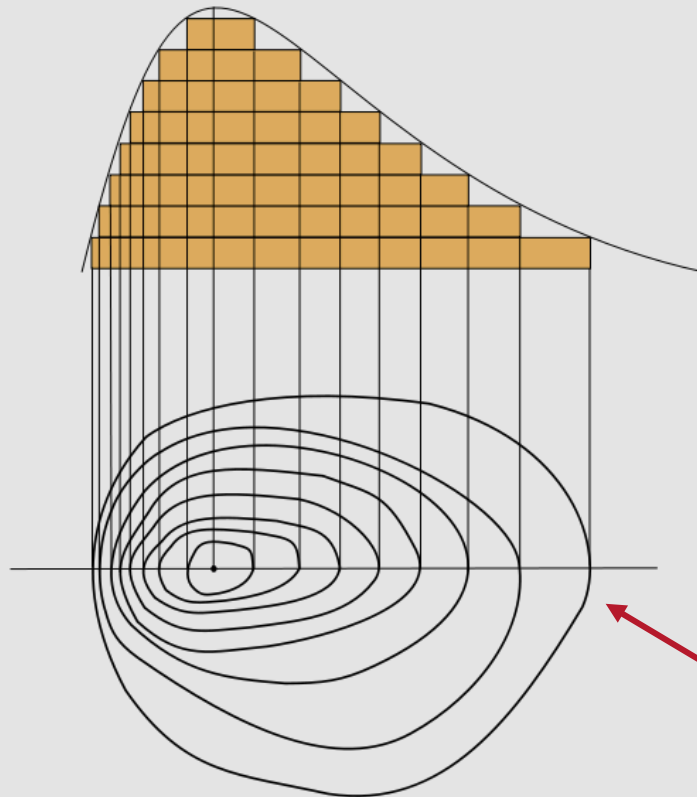
Contour line

A contour line is a decision rule.

Just try to imagine that the associated heights are incorrect.

How 'certain' would you be?

decision rule

Thursday, October 20, 2011

Public information from WADA

☐ World anti-doping agency (WADA): "Unanimous opinion of the panel that there is no known reasonable explanation for the blood profile information of this Athlete other than the use of a Prohibited Substance or Prohibited Method."

☐ Alternative view: Once flawed statistics 'reveals' that the fluctuation(s) cannot be explained by chance, 'a panel' performs 'a review'. This ad-hoc two-stage procedure opens the door for confirmation bias, aka tunnel vision.

☐ Recall widespread EPO abuse in the past.

Thursday, October 20, 2011

…no known reasonable explanation…

- A classical logical fallacy: argument from ignorance (argumentum ad ignorantiam), the so-called negative proof.

- The Pechstein case illustrates that it is extremely difficult to counter negative proof.

12

Suspicion or valid proof?

□ One has a smoking gun, namely suspicious blood values.

□ Next, the burden of proof is shifted to the athlete, without actually providing a proof of doping. After all, the athlete is provided the opportunity to prove his/her innocence.

□ Undoubtedly, this will often work as intended. However, if the athlete is innocent, the 'abnormal' blood values are caused by chance, and the procedure amounts to "torturing the data until they confess (for the athlete)".

□ In conclusion, there's no defense for an innocent person.

Public information from the cycling union

What is a biological passport?

A biological passport is an individual, electronic record for each rider, in which the results of all doping tests over a period of time are collated.

Can the haematological profile be used to open a "doping case"?

The scientific assessment of a rider's profile applies similar principles to those used in forensic medical science to determine the likelihood of guilt.

Misleading: '…applies similar…'

Source (05-10-2011): http://www.uci.ch > Anti-doping > Questions/Answers

Thursday, October 20, 2011

# 3. Some problems with statistics

□ There are various problems, some of them are fundamental and difficult to repair in practice.

□ Limit to a few to illustrate:

- decisive flaws in the Pechstein case;
- the connection with the case of Lucia de Berk: misleading focus on 'abnormal' - the wrong probability!

□ Throughout, keep in mind that these problems tend to increase the risk of a false-positive.

Three probabilities required
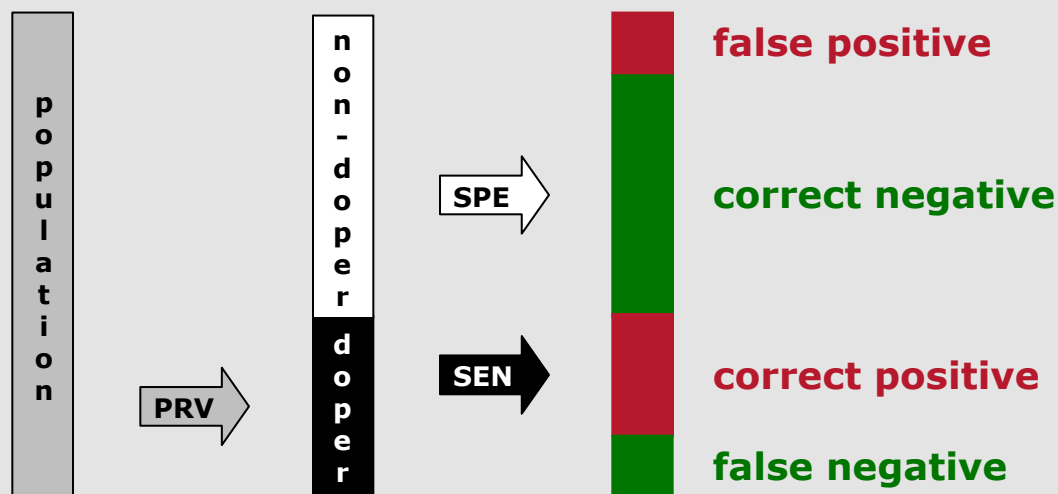
PRV = prevalence = fraction of dopers in population
SPE = specificity = probability of correct-negative
SEN = sensitivity = probability of correct-positive

How to estimate these probabilities?

☐ Assume that representative 'clean' volunteers are available, so that specificity is under control.

☐ When doping habits are known, sensitivity can be reliably estimated by supplying doping to volunteers.

☐ With respect to estimation of prevalence, the biostatistician Don Berry notes (*Chance*, 2004, **17**: 8-11): "Prevalence of disease is relatively easy to estimate, depending on the patient population. Prevalence of substance abuse is not. There is an inevitable subjective aspect of assigning a prior probability. A hearing board made up of an athlete's peers is especially appropriate for making such assignments."

17

Numerical example

PRV	=	prevalence	=	fraction of dopers in population	=	10 %
SPE	=	specificity	=	probability of correct-negative	=	99.9 %
SEN	=	sensitivity	=	probability of correct-positive	=	5 %

☐ Specificity is good. Few clean athletes have 'abnormal' values.

☐ Sensitivity is so, so. However, with higher sensitivity the abuse quickly stops (repeated testing).

☐ Take a population of 10,000 athletes:
- 0.1% x 9000 = 9   false-positives
- 5% x 1000 = 50 correct-positives

☐ Probability of guilt: 50 / 59 = 85%. By far not good enough.

18

Specificity-based decision rule (1/4)

adverse result = outlier

Results Male Athlete

'abnormal'

decision rule

99.9% 'certain'

b l o o d v a l u e

Source: Anne Gripper (cycling union), ANADO Workshop (2008)

Thursday, October 20, 2011
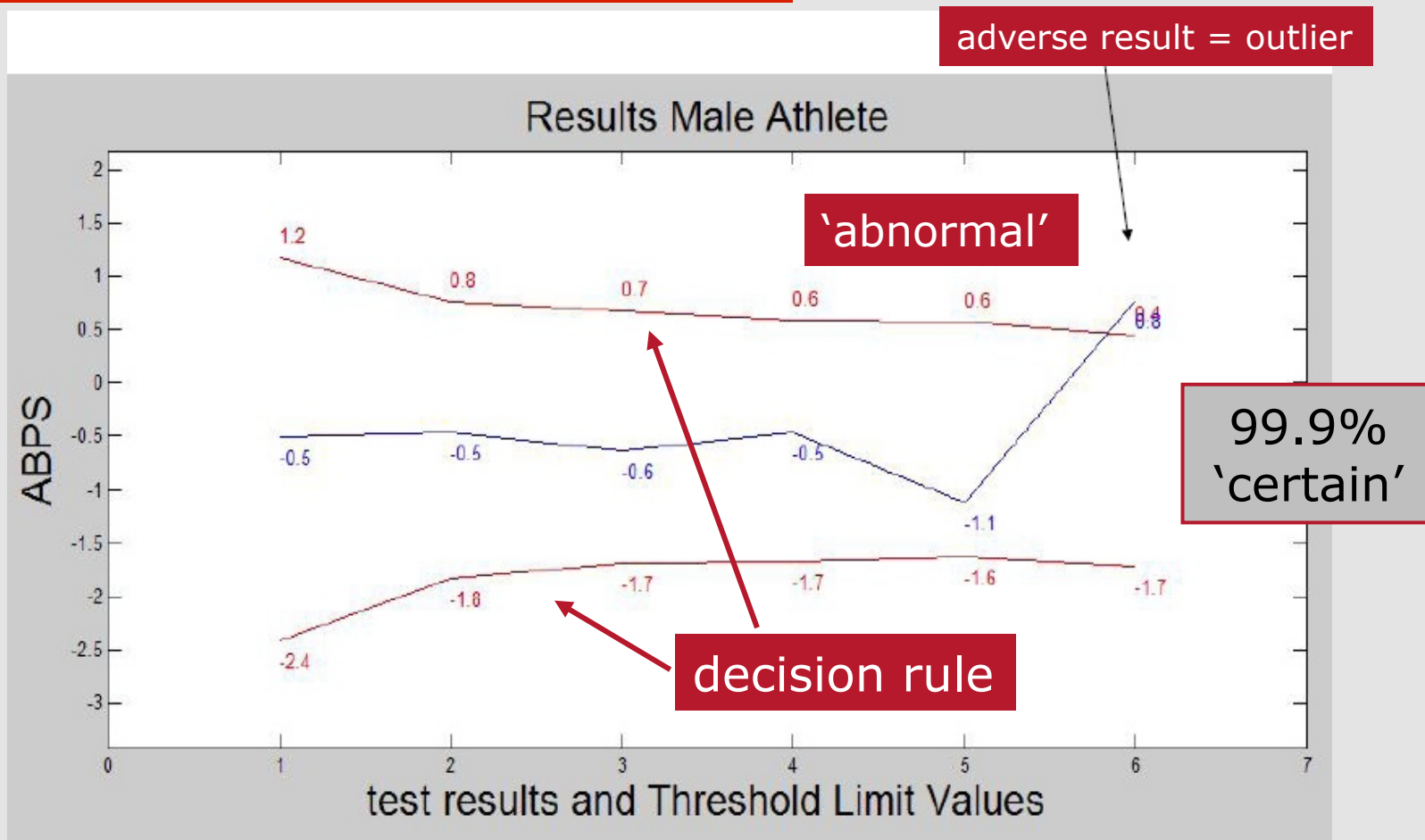
Specificity-based decision rule (2/4)

- ☐ Take a group of 1000 'clean' athletes: only one (1) is automatically picked out and prosecuted, etc.

- ☐ This procedure amounts to a false-positive risk of 1 in 1000 for each athlete in the group. That's irrelevant!

- ☐ The false-positive risk is 100% for the unlucky athlete who happens to be selected. That counts!

- ☐ However, the alleged offender incorrectly gets the label of 99.9% 'certainty' of being doped.

20

Thursday, October 20, 2011

Specificity-based decision rule (3/4)

☐ Plausibility accepted for the model:

99.9 % or 999/1000
Potential false positive: 1/1000

☐ After experts review and hearing the risk of false positive is closed to zero

Alain Garnier

Source: Alain Garnier, WADA Media Symposium (2008)

Specificity-based decision rule (4/4)

- E-mail of the Anti-Doping Authority Netherlands, dated October 14, 2008:

- "ik zag bv ook dat Sottas de term 'vals positieve' verkeerd definieert"

- "I saw e.g. also that Sottas incorrectly defines the term false-positive"

- A minor detail, hardly a reason to undertake action

22

Prosecutor's fallacy

□ A specificity-based decision rule invites for prosecutor's fallacy. The underlying fallacious reasoning runs as follows:

1. There is a probability of 0.1% (say) that this evidence ('abnormal' value) is found if the suspect is innocent ('clean');
2. Therefore (?), the probability that the suspect is innocent should, likewise (?), be 0.1%. It is immediate (?) that the likelihood of guilt is 99.9%.

□ It is essential to balance specificity against the other probabilities, namely sensitivity and prevalence.

Thursday, October 20, 2011

Low prevalence in speed skating!

☐ Harm Kuipers (skating union ISU):

"The ISU has tested 1650 speed skaters since 2000. Urine tests have led to three (3) suspensions. One can state in general that doping is not a big issue in speed skating."

☐ Of course, 1:550 is a naïve estimate of prevalence – too small.

☐ However, skating is not comparable at all to professional cycling.

☐ This situation favors committing prosecutor's fallacy.

24

Source: A. ten Broeke, 'De uitschieter van Pechstein', *Spits,* 15 januari 2010

Distributional assumption

□ A normal distribution is assumed for 'abnormal' values.

□ Richard Gill (Arnout Jaspers, *nwt*, 2010, **2**: 34-35): a dogma from the 19th century, proven not to work for medical issues.

□ Cliff Spiegelman (Ewen Callaway, *Nature*, 2011, **475**: 283-285): "The problem, says Speigelman, is that biological measurements are chock full of outliers — far more than would be predicted by a normal distribution. Proponents of the passports are "presenting themselves as more accurate than they really are", he says, and he estimates that the false-positive rate of the passport could be off by a factor of 10 or even 100."

25

Model re-use is ignored

- Obviously, the model is to be used more than once. However, this not accounted for (!) in the calculations.

- Say, the 'certainty' (=specificity) is 99% if the model is used once. Then the 'certainty' is 99%×99%≈98% if the model is used twice, etc.

- Conclusion: the correct 'normal' range follows from a so-called tolerance interval instead of the current confidence interval.

26

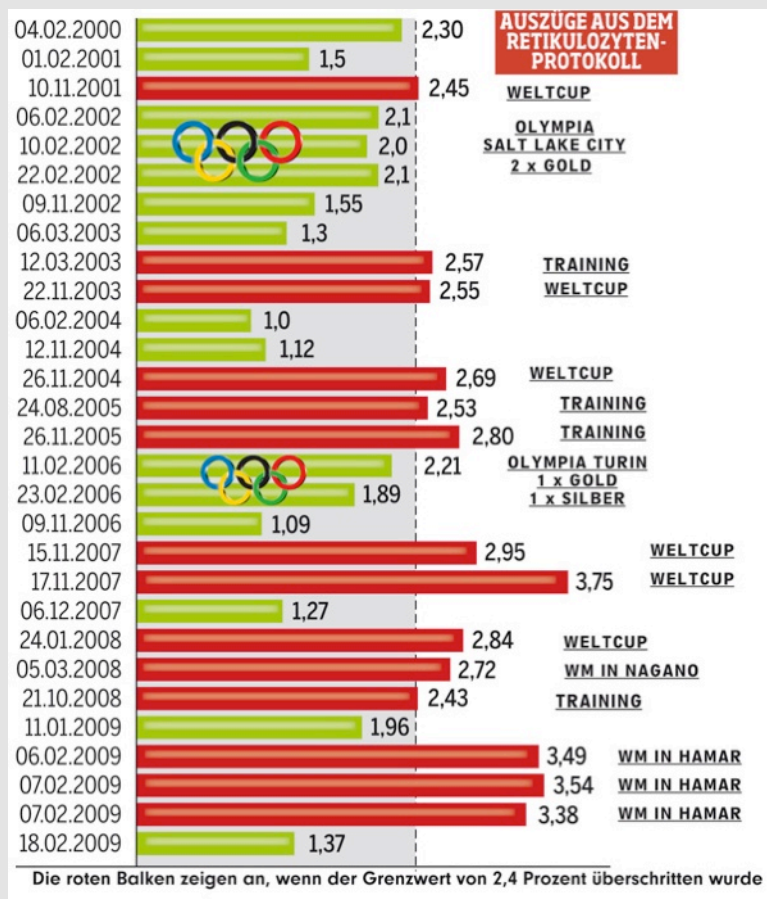Thursday, October 20, 2011

4. The ongoing case of Claudia Pechstein

despite

□ Five-time Olympic gold medalist.

□ The first ever to be convicted ~~owing to~~ her biological passport.

Young red blood cells (%RET)

Date	Value	Event
04.02.2000	2,30	
01.02.2001	1,5	
10.11.2001	2,45	WELTCUP
06.02.2002	2,1	OLYMPIA SALT LAKE CITY 2 x GOLD
10.02.2002	2,0	
22.02.2002	2,1	
09.11.2002	1,55	
06.03.2003	1,3	
12.03.2003	2,57	TRAINING
22.11.2003	2,55	WELTCUP
06.02.2004	1,0	
12.11.2004	1,12	
26.11.2004	2,69	WELTCUP
24.08.2005	2,53	TRAINING
26.11.2005	2,80	TRAINING
11.02.2006	2,21	OLYMPIA TURIN 1 x GOLD 1 x SILBER
23.02.2006	1,89	
09.11.2006	1,09	
15.11.2007	2,95	WELTCUP
17.11.2007	3,75	WELTCUP
06.12.2007	1,27	
24.01.2008	2,84	WELTCUP
05.03.2008	2,72	WM IN NAGANO
21.10.2008	2,43	TRAINING
11.01.2009	1,96	
06.02.2009	3,49	WM IN HAMAR
07.02.2009	3,54	WM IN HAMAR
07.02.2009	3,38	WM IN HAMAR
18.02.2009	1,37	

AUSZÜGE AUS DEM RETIKULOZYTEN-PROTOKOLL

Die roten Balken zeigen an, wenn der Grenzwert von 2,4 Prozent überschritten wurde

Observations:

1. Average is 2% instead of the 'normal' 1%;

2. Often high level during events.

Remarks:

1. Because of the 'abnormal' average, Mrs. Pechstein had always been under suspicion before she could be barred from competition at Hamar;

2. These values were not shared with Mrs. Pechstein.

The value of the indirect evidence??

1. Jan 1, 2009: WADA enables prosecution on the basis of ('abnormal') blood values only.

2. Feb 6-7, 2009: So-called 'abnormal' blood values observed in Hamar, NOR. Mrs. Pechstein barred from competition.

3. Feb 12, 2009: Data sent to Lausanne laboratory for further scrutiny. That fact alone proves they improvised in Hamar.

4. Nov 25, 2009: Sports tribunal (CAS) award published. It contains a simplistic assessment of 'abnormality '.

5. Dec 2, 2009: WADA operating guidelines approved. Mrs. Pechstein would still have skated under these guidelines (doing the wrong thing right!).

6. 2010: Various requests sent to Lausanne laboratory to provide details about their assessment of the data. No reply.

29

WADA operating guidelines

□ The specificity-based decision rule:

"A profile in which the Adaptive Model has identified the Hb or Off-hr score abnormal with a 99.9% probability or more shall be reviewed by a panel of three experts."

(…)

□ "In the panel's unanimous opinion, absent a satisfactory explanation from the *Athlete*, it is highly likely that the *Athlete* has used a *Prohibited Substance* or *Prohibited Method*;"

□ Two parameters must be considered:
- Hb = hemoglobine;
- Off-hr = Hb − 60 √%RET.

30

Thursday, October 20, 2011

Confirmation bias at work

- Mrs. Pechstein was prosecuted on the basis of her young red blood cells only, which is a school example of confirmation bias
  - Wikipedia: "Confirmation bias is the tendency for people to only seek out information that conforms to their pre-existing view points, and subsequently ignore information that goes against them. It is a type of cognitive bias and a form of selection bias toward confirmation of the hypothesis under study. Avoiding confirmation bias is an important part of rationalism and in science in general."
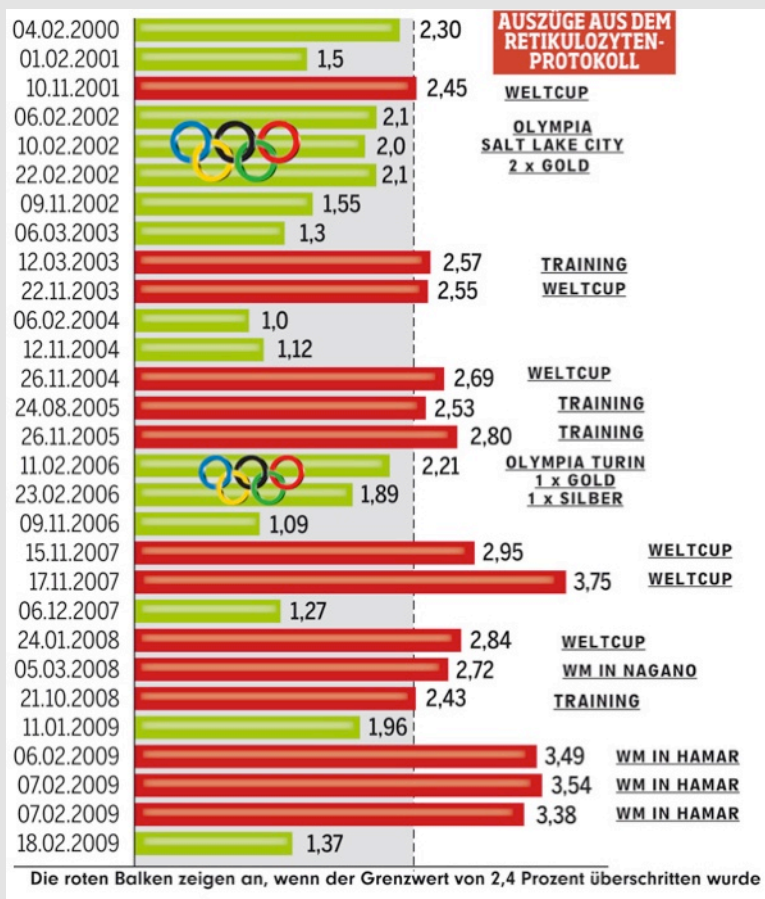
- The level of hemoglobin (Hb) was constant and not particularly high during the entire period of 10 years.
- Reason for not considering Hb: Hb can be manipulated by diluting the blood, whereas the percentage reticulocytes is robust against manipulation. It is relative to all red blood cells.

31

Evidence or counter evidence??

☐ If anything, a high level of reticulocytes during major events is counter evidence of blood doping.

☐ The timing is way off.

☐ How can the skating union (ISU) see evidence here? Tunnel vision?

32

Thursday, October 20, 2011

WADA: reasonable explanation??

Die roten Balken zeigen an, wenn der Grenzwert von 2,4 Prozent überschritten wurde
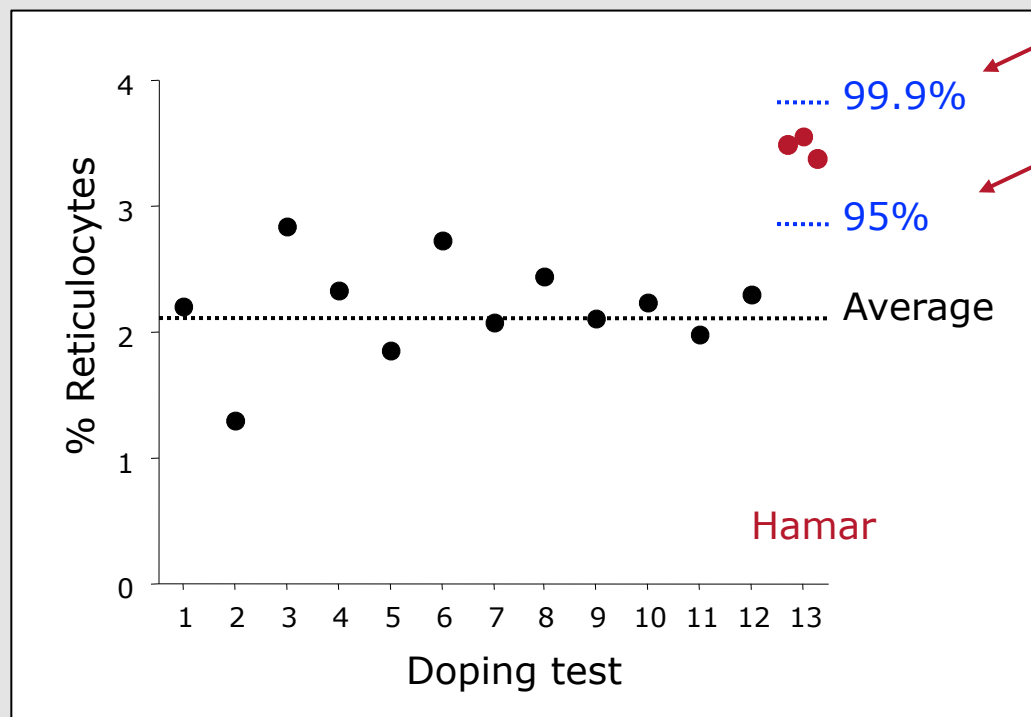
- □ Recall: often high level of young red blood cells during major events.

- □ Medical experts:
  - → stress

- □ Non-medical anti-doping researchers like Werner Franke:
  - → unknown method or drug

Wrong threshold for wrong probability!

WADA guideline

ISU ad-hockery

Three tests within two days: being extremely careful or overlooking the Texas sharpshooter fallacy??

Older data discarded owing to technicalities.

Implication of this observation alone

From paragraph 117 in the sports tribunal (CAS) award:

"even in cases of adverse analytical findings, departures from WADA international standards do not invalidate per se the analytical results, as long as the anti-doping organisation establishes that such departure did not cause the adverse analytical finding"

Apparently, this observation remains unrecognized and/or unappreciated.

WADA, May 13, 2010

Mr Faber,

As usual, serious misunderstanding on your part (which I already explained to you in a previous mail in December 2009): Mrs Pechstein's case was not reviewed under the Athlete Passport model as developed by WADA, but under the ISU longitudinal model. Not the same rules and not the same science in support.

Sincerely,

Olivier Rabin
Director, sciences

36

Response to complaints

Harm Kuipers (ISU)

"Statisticians have looked at Faber's complaints. They were not impressed."

Olivier de Hon (Anti-Doping Autority Netherlands)

"They have applied a certainty of 99.99% in this case."

Pierre-Edouard Sottas (Lausanne laboratory)

"it is not possible to analyse the blood data of Mrs Pechstein with the model I developed, simply because the collection, transport and analysis of the blood samples did not follow the corresponding WADA protocols"

A ticking time bomb since her ban ended

☐ Research has revealed that the elevated level of reticulocytes is caused by a blood disorder, inherited from the father.

☐ This result should have no legal implications since the elevated profile has been accepted by the the sports tribunal CAS:

- ■ "(…) the Panel is of the opinion that, to the benefit of the Athlete, the mean value of 2.1 might be safely taken into consideration as a basis for comparison of the Athlete's %retics values recorded in Hamar on 6 and 7 February 2009."

☐ As noted earlier, outliers are less rare than predicted from a normal distribution. In March (Inzell), Mrs. Pechstein produced new outliers, which was to be expected. What to do now??

38

5. Concluding remarks

- Before the trial:
  - It cannot be denied that the elevated level of young red blood cells has led to a highly persistent suspicion
  - If WADA's guidelines had been applied in Hamar, Mrs. Pechstein would have continued to skate

- During the trial:
  - The panel approved (!) the elevated level, to be able to flag the values attained in Hamar as 'abnormal'
  - Subsequently, Mrs. Pechstein was convicted owing to a simplistic ad-hoc assessment of 'abnormality'

- After the trial:
  - New 'abnormal' values have been observed, thus confronting the skating union ISU with an unsolvable problem

39