

LESERBRIEF

zu dem Editorial von Steinacker JM: „Good or bad? Fairness and Dilemma in Anti-Doping Fight“ *Dtsch Z Sportmed* 61 (2010) 3

Failing intuition and lying statistics have condemned Claudia Pechstein

Sir:

Don't we all experience rare events in daily life, sometimes extremely rare indeed? So, how rare is Mrs. Pechstein's reticulocyte value (3.5%) observed at the World Championships at Hamar, February 7-8, 2009? Considering both her generally elevated level (2.1% over the last 17 samples before the Hamar event vs. a 'normal' average of ca. 1%) and her intra-individual variation (standard deviation of 0.42%), straightforward calculations (not shown for brevity) lead to odds that are a little over 1:1000. Is that really sufficiently rare to prosecute and subsequently convict her for blood doping? A crash course in statistics follows, as well as a short description of two famous miscarriages of justice to illustrate that, in my opinion, history merely repeats itself.

A crash course in statistics

Statistics is a rigorous methodological discipline that finds application in various fields of human endeavor. It even helps answering rather exotic questions like "What day of the week has the least rainfall?". In the Netherlands, data about rainfall have been compiled since 1910. Just recently, it has been published that over the last century (i.e. 1910-2009), Saturday had the least rainfall with an average of 2.057mm, while Thursday was wettest with an average of 2.240mm. The overall weekly average was 2.172mm. Thus, what to infer from these daily fluctuations? The correct (intuitive) answer is: nothing. Not surprisingly, proper statistical evaluation of the data reveals that the differences with respect to the overall weekly average can indeed be ignored for practical purposes: they are not 'statistically significant'. This type of information is extremely useful. It arms a decision maker with an objective safeguard against 'hineininterpretieren' (and subsequent unwarranted action), for lack of a better term in English. Let's now turn attention to a didactic example that has direct relevance to the proper assessment of the data measured for Mrs. Pechstein.

Consider a lottery that sells 1000 tickets (say), only one of which is winning. Now one might ask oneself: given the obvious odds of 1:999 of winning, what is the chance that the 'lucky' winner has cheated to win? Surely, one must agree that the available information is grossly incomplete. Consequently, the correct (intuitive) answer immediately follows as: there's no way of knowing the truth. In particular, it's not a 99.9% chance of guilt. Therefore, without additional incriminating information, lottery winners receive their prize, instead of a visit by the police. In a strict sense, winning a lottery is indirect evidence of cheating (since the large majority of honest players loses over and over again), but the evidence is by far not strong enough by itself to prosecute. Let's now turn attention to the evidentiary weight of a single rare reticulocyte value to detect blood doping.

Consider 1000 athletes (say) and measure reticulocyte values once to detect blood doping. N.B. Exactly the same reasoning applies to a group of 100 athletes for which 10 reticulocyte values are measured. (The latter situation more resembles the data base of female speed skaters.) Measuring the largest fluctuation among the

1000 athletes is perceived as a rare event, just like drawing the winning ticket in a lottery. Therefore, in straightforward analogy with the 'theoretical' lottery example, the largest fluctuation constitutes indirect evidence of cheating. However, one should be extremely careful when bothering this 'unlucky' athlete: the available information is grossly incomplete. It is exactly this methodological flaw that has been clearly overlooked in the case of Mrs. Pechstein. Viewed differently: intuition has failed. Statistically speaking, these rare events simply must occur with high probability due to multiple testing (multiple athletes, multiple tests per athlete), regardless any administration of blood doping.

On a fully abstract level, logic dictates that one cannot generate a hypothesis (guilty of cheating) and confirm it, using the same data – the Hamar fluctuation, winning a lottery, etc. One needs additional incriminating information to confirm the hypothesis / to complete the proof. Moreover, without additional incriminating information probabilities do not make sense. At this point, I urge the reader to really take the time to digest the troublesome fact that the prosecution has not reported a single relevant probability in the case of Mrs. Pechstein.

In summary, there has never been a proof of doping, only an unconfirmed suspicion. Unfortunately, statistics is a rather abstract methodological discipline and therefore poorly represented in a court of law. The case of Mrs. Pechstein appears to be entirely obscured by medical details that bear the attractive feature of being tangible and concrete, although they might be trivial for the correct outcome of the trial.

Two famous miscarriages of justice

Quite recently (April 14, 2010), a notorious criminal case was concluded in the Netherlands with the complete exoneration of Lucia de Berk, previously known as the angel of death. Mrs. de Berk was convicted as a serial killer to a life time sentence in jail because of indirect evidence incorrectly assessed. N.B. The initial odds against her were even much more extreme [1]:

„2003 wurde die Krankenschwester wegen Mordes an sieben Patienten, darunter dem Baby, zu lebenslanger Haft verurteilt. In dem Prozess kam unter anderem ein Statistik-Experte zu Wort, der die Wahrscheinlichkeit, dass De Berk zufällig bei allen verdächtigen Todesfällen Dienst hatte, mit eins zu 342 Millionen angab.“

In many ways, the current case is a copy of that one. Likewise, it very much resembles yet another famous miscarriage of justice: the case of Sally Clark. Now again, possibly trivial details distract from crucial methodological issues.

Both criminal cases have been described in an article published in *Nature* [2]. The two statistics professors interviewed for that article support me in the claim that an abuse of statistics played a decisive role in the case of Mrs. Pechstein [3].

In Lucia de Berk's case, it eventually took the responsible magistrates seven (7) years to recognize and admit that it was the lying statistics that had condemned her. Mrs. de Berk has been promised swift and full compensation for the six years she spent in prison. Mrs. Clark never recovered from the trial and subsequent imprisonment, and died soon after she was released.

Let's hope that lessons are finally learned from these sad cases so that abstract statistical arguments no longer fall on

stony ground. Lottery winners do not automatically draw the attention of the police, and rightly so.

The guidelines of the World Anti-Doping Agency (WADA)

It has been argued in the media that Mrs. Pechstein's case should have been reviewed under WADA's operating guidelines [4]. Calculations (not shown for brevity) show that this could have affected the outcome. This observation has profound legal implications. From paragraph 117 in the CAS award [5]:

“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”

I conclude by noting that WADA's operating guidelines are statistically flawed as well: the false-positive risk is systematically underestimated [6-8]. One can imagine that for a basic scientist like myself, it is extremely embarrassing to witness that 'doing the wrong thing right' as early as the Hamar event, would have avoided the case to develop as it did – with unnecessary prolonged litigation.

Literature

1. http://www.nzz.ch/nachrichten/panorama/hollaendischer_todesengel_unschuldig_1.5446938.html; Consulted July 28, 2010.
2. **BUCHANAN M:** Conviction by numbers. *Nature* 445 (2007) 254-255.
3. http://www.nrc.nl/sport/article2430747.ece/Statisticus_laakt_zaak-Pechstein; Consulted July 28, 2010.
4. http://www.wada-ama.org/Documents/Science_Medicine/Athlete_Biological_Passport/WADA_AthletePassport_OperatingGuidelines_FINAL_EN.pdf; Consulted July 28, 2010.
5. <http://www.tas-cas.org/d2wfiles/document/3802/5048/0/FINAL%20AWARD%20PECHSTEIN.pdf>; Consulted July 28, 2010.
6. **FABER K, SJERPS M:** Anti-doping researchers should conform to certain statistical standards from forensic science. *Sci Justice* 49 (2009) 214-215.
7. **FABER NM, VANDEGINSTE BGM:** Flawed science 'legalized' in the fight against doping: the example of the biological passport. *Accred Qual Assur* 15 (2010) 373-374.
8. **SPIEGELMAN C:** Senior statisticians need to be involved. *Accred Qual Assur* 15 (2010) in press.

Correspondence:

Klaas Faber, Ph. D.

Chemometry Consultancy

Goudenregenstraat 6

6573 XN Beek-Ubbergen

The Netherlands

E-Mail: nmf@chemometry.com