Letter to the Editor

On the calculation of decision limits in doping control

Nicolaas (Klaas) M. Faber[1,*] and Ricard Boqué[2]

[1]Chemometry Consultancy, Rubensstraat 7, 6717 VD Ede, The Netherlands

[2]Department of Analytical Chemistry and Organic Chemistry, Rovira i Virgili University, 43007 Tarragona, Spain

[*] Communicating author:
e-mail: nmf@chemometry.com
Tel.: +31-318-641985
Fax: +31-318-642150

Sir,

In doping control in sport, measurements of analyte level are classified as either negative or positive by comparison with a decision limit. In case of a negative declaration, no action follows. A positive declaration, however, implies serious consequences for the athlete in question. It therefore stands to reason that the doping laboratory must deploy a decision limit that ensures a sufficiently small risk of a *false positive* declaration (α). Moreover, the correct calculation of decision limits should not be a subject of dispute. However, two opposing views have been put forth in this Journal concerning the correct calculation of decision limits in doping research [1,2]. It is important to note that these opposing views may lead to vastly different estimates when measurement uncertainties depend on the level of the analyte of interest, which is often the case. For example, values discrepant by a factor of five have been reported for 19-norandrosterone (males) [1,2]. The purpose of this Letter is to clarify that this issue has been conveniently settled in the statistics-oriented literature. In other words, the correct calculation of decision limits should indeed not be a subject of dispute at all.

How to control the risk of false positives?

We insist that a correct decision limit should ensure a sufficiently small risk of a *false positive* declaration (α). For this purpose, rigorous methodology has been developed from the theory of hypothesis testing [3]. This methodology is now recommended by the International Union of Pure and Applied Chemistry (IUPAC) [4].[1]

Consider the case that the analyte of interest should not exceed a certain threshold (L).[2] Further assume that the measurement uncertainty is proportional to the level of the analyte of interest, since this is also assumed in [1,2]. Denote this relative uncertainty as u_{R}. Then the correct decision limit (X_M) follows by considering the uncertainty in the measurement value at the threshold level, since this level is the highest level that complies with the null-hypothesis, namely the athlete in question is 'clean':

$$X_M = L + k \cdot u_{\mathrm{R}} \cdot L = L + U \qquad (1)$$

[1] Harmonization in this respect of the International Organization for Standardization (ISO) and IUPAC has been thoroughly documented in [5]. For a recent review of practical examples, see [6].

[2] The notation is slightly adapted from [7].

where the value of the coverage factor k depends on the desired level of confidence, and $U = k \cdot u_R \cdot L$ denotes the expanded uncertainty. This calculation method, which is favored in [2], is depicted in Figure 1. Throughout, the value of 1% for α is taken for illustrative purpose only. Figure 1 constitutes a slightly adapted version of Figure 2 in [7].

The problem with the alternative calculation method

The alternative calculation method, which is favored in [1], is depicted in Figure 2. (This is a slightly adapted version of Figure 1 in [7].) Here the decision limit is directly related to the left tail of the distribution of true positives. In other words, this calculation method controls the risk of false negatives (β). This calculation method is, however, motivated in [1] as follows: "For a quantitative measurement, the decision limit is defined as (article 6.2 of [8]): "If a permitted limit has been established for a substance, the decision limit is the concentration above which it can be decided with a statistical certainty of 1 - α that the permitted value has been truly exceeded." This definition can be compatible with the one-sided *t*-test as given by WADA [9], and the scheme given in ILAC G8 [10]."[3] It is observed, however, that α is replaced by β! King [7] also argues that Figure 2 is consistent with the WADA [9][4] and ILAC G8 [10] strategies and adds a reference to ISO [11]. King [7] further increases the confusion by stating "The approach used in Ref. [1] to be sound, but the value of the coverage factor used seems inappropriately high (in the absence of any justification)".

The reasoning behind the alternative calculation method is straightforward, but flawed: there is a 1% probability of a false negative result hence "we can be confident at the 99% level that the result is positive" [7]. However, to arrive at these numbers one has to assume a *distribution of true positives*.[5] This assumption has undesirable implications. Consider the case where the measurement value only slightly exceeds the decision limit. Then, under the alternative strategy, *any* indication (even much less than the conventional values of 95% or 99% certainty) has become a solid proof because there are *no false*

[3] References [8], [9] and [10] correspond to References [28], [4] and [3] in [1].
[4] This reference is updated to Version 3 (June 2003).
[5] Clearly, a 1% probability of a false negative result must be supplemented by 99% true positives.

positives.[6] The uncertainty in the measurement value has effectively disappeared at the decision stage.[7] To avoid this kind of problems, conventional hypothesis testing centers on the null-hypothesis. Action is taken only when the measurement value sufficiently supports rejection of the null-hypothesis. Specifically, an athlete is considered to be positive when the measurement value is unlikely to be obtained for a 'clean' sample. Subsequent steps may be taken to verify the positive as true or false. This strategy is statistically sound, hence fair to all parties.

Concluding remarks

It is tempting to speculate on the reasons for adopting the fallacious alternative calculation method.[8] One reason could be that both methods lead to the same decision limit when measurement uncertainty is not level-dependent (and $\alpha = \beta$). However, a procedure that leads to grave consequences in case of a false positive declaration must simply guard against false positive declarations! It follows that one should examine the right tail probability of the distribution of true negatives (null-hypothesis).

[6] If this result is obtained for the A-sample, analysis of the B-sample is no longer necessary!
[7] Uncertainty does of course play a role in the calculation of the decision limit.
[8] In a classical paper by Bland and Altman [12] (11,194 citations on the ISI Web of Science, 7 July 2006), a widespread misconception is aptly discussed as follows: "Why has a totally inappropriate method, the correlation coefficient, become almost universally used for this purpose? Two processes may be at work here – namely, pattern recognition and imitation. (…) Journals could help to rectify this error by returning for reanalysis papers which use incorrect statistical techniques. This may be a slow process. Referees, inspecting papers in which two methods of measurement have been compared, sometimes complain if no correlation coefficients are provided, even when the reasons for not doing so are given."

References

1. van der Veen AMH (2003) Accred Qual Assur 8:334-339
2. Van Eenoo P, Delbeke FT (2003) Accred Qual Assur 8:477-479
3. Currie LA (1968) Anal Chem 40: 586-593
4. Currie LA (1995) Pure Appl Chem 67:1699–1723
5. Currie LA (1997) Chemom Intell Lab Syst 37:151-181
6. Olivieri AC, Faber NM, Ferré J, Boqué R, Kalivas JH, Mark H (2006) Pure Appl Chem 78:633-661
7. King B (2004) Accred Qual Assur 9:369-373
8. European Commission (2002) Council Decision of 12 August 2002, implementing Council Directive 96/23/EC concerning the performance of analytical methods and interpretation of results, 2002/657/EC, Official Journal of the European Communities, L221/8 EN, 17.08.2002
9. World Anti-Doping Agency WADA (2002) Laboratory accreditation requirements and operating standards, Version 1.0, 10 November 2002. WADA: http://www.wada-ama.org/
10. International Laboratory Accreditation Cooperation (ILAC) G8 (1996) Guidelines on the assessment and reporting of compliance with specification. ILAC: http://www.ilac.org/
11. ISO/DIS 10576-1 (2000) Statistical methods—guidelines for the evaluation of conformity with specified requirements. Part 1. General principles. ISO, Geneva, Switzerland
12. Bland JM, Altman DG (1986) Lancet 327:307-310

Fig. 1 Distribution of measurement values when the analyte of interest is present at the threshold level (*L*), i.e. true negative results. The detection decision is taken at a level (X_M) that is sufficiently higher than the threshold to ensure that the risk of a false positive declaration (α) will not exceed a predetermined value, e.g. 1%. This procedure is essentially correct, since controlling the risk of false positive declarations is the key responsibility of the doping laboratory. For the calculation of *U*, see Equation (1).

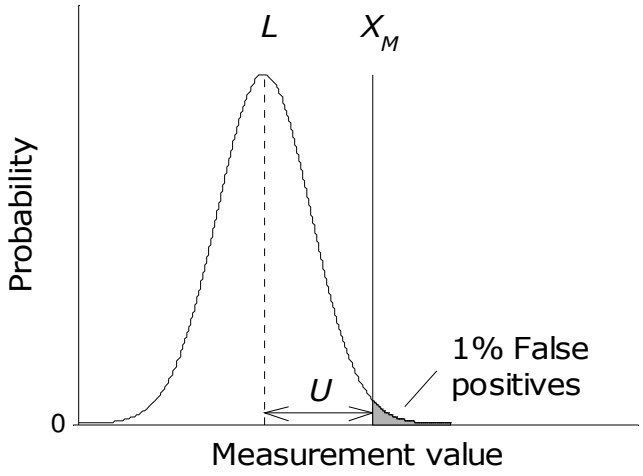

Fig. 2 Distribution of measurement values when the analyte of interest is present beyond the threshold level (L), i.e. true positive results. The detection decision is taken at a level (X_M) that is sufficiently higher than the threshold to ensure that the risk of a false negative declaration (β) will not exceed a predetermined value, e.g. 1%. This procedure does not make sense, since controlling the risk of false negative declarations is irrelevant in the context of doping control laboratories. For the calculation of U, see [1].