Journal Title
XX(X):2–13

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Exact statistical calculation of the uncertainty term in the decision limits based on the GH-2000 score for growth hormone misuse detection (doping)

Dankmar Böhning,[1] Wei Liu,[1] Richard I.G. Holt,[2] Walailuck Böhning,[2] Nishan Guha,[3] Peter Sönksen,[2] David Cowan,[4] and Tianyi Liang[1]

Abstract

The GH-2000 score has been developed as a powerful and unique technique for the detection of growth hormone misuse by sportsmen and women. The score depends upon the measurement of two growth hormone (GH) sensitive markers, insulin-like growth factor-I (IGF-I) and the amino-terminal pro-peptide of type III collagen (P-III-NP). It also includes a term to adjust for the age of the athlete. Decision limits for the GH-2000 score have been developed and are incorporated into the guidelines of the World Anti-Doping Agency (WADA). These decision limits are derived by setting a 1 in 10,000 false-positive rate rule. As these decision limits are estimated from samples of GH-2000 scores they carry uncertainty. In previous work, this uncertainty has been addressed by establishing an upper 95% confidence interval for the true decision limits based on a normal approximation which has been shown to be appropriate if sample sizes are large (such as 1000 and above). Here we show that these approximations, whether reasonable or not, can be entirely avoided by developing an upper 95% confidence interval for the true decision limits using an approach based upon the t-distribution. While there are considerable differences for smaller sample sizes, these become negligible when the sample size is large such as 1,000 and above.

Keywords

Growth hormone misuse detection; GH-2000 score; Decision Limits; Tolerance Limits

Prepared using sagej.cls *[Version: 2015/06/09 v1.01]*

1 Introduction

Growth hormone is a powerful anabolic agent of considerable therapeutic value but also misused in sport for its anabolic and lipolytic properties.[1] In order to preserve the fairness of competition, its use is prohibited by the World Anti-Doping Agency[2] and there is a need for methods to detect its misuse. Two methods are presently available and approved by the World Anti-Doping Agency (WADA); the isoform test developed by Bidlingmaier et al.[4] (see also WADA[2,3]) and the GH-2000 biomarker test developed by the GH-2000 and GH-2004 projects[5]. The latter method depends upon the measurement of two growth hormone (GH) sensitive markers, insulin-like growth factor-I (IGF-I) and the amino-terminal pro-peptide of type III collagen (P-III-NP), both of which rise in response to exogenous GH administration[8,9]. The measured concentrations of the biomarkers are combined in sex-specific and age-adjusted discriminant functions, which allow for the calculation of a score (the GH-2000 score) on which basis the compliance of the samples' analytical result is determined.

The paper is organized as follows. In section 2 further details on the GH-2000 score construction is given as well as the background of decision limit construction to detect GH misuse is reviewed. A crucial element in the decision limit construction is how uncertainty is incorporated. In section 3, we present a new and alternative way of incorporating uncertainty in an exact statistical way, given a desired false-positive rate of 1 in 10,000. The paper closes with a discussion putting these findings in perspective.

[1] Mathematical Sciences & Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, SO17 1BJ, UK
[2] Human Development and Health Academic Unit, Faculty of Medicine, The Institute of Developmental Sciences (IDS Building), Southampton General Hospital, University of Southampton, Southampton, SO16 6YD, UK
[3] Chemical Pathology and Metabolic Medicine Department of Clinical Biochemistry, John Radcliffe Hospital, Headley Way, Headington Oxford OX3 9DU, UK
[4] Drug Control Centre, Department of Pharmacy & Forensic Science, King's College London, 150 Stamford Street, London SE1 9NH, UK

Corresponding author:
Dankmar Böhning, Mathematical Sciences and Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, SO17 1BJ, UK.
Email: D.A.Bohning@soton.ac.uk

Prepared using sagej.cls

2 Background and state-of-the-art approach for constructing the decision limits for growth hormone misuse

The GH-2000 score has been developed in Powrie et al.[10], Erotokritou-Mulligan et al.[11] and Holt et al.[5] It has the theoretical or model form

$$\beta_0 + \beta_1 \log \text{IGF-I} + \beta_2 \log \text{P-III-NP} + \beta_4 \frac{1}{\text{age}} \tag{1}$$

where the known coefficients β_i's have different values for male and female athletes. IGF-I and P-III-NP are generic terms for the two serum proteins to be measured. This measurement process is done by choosing a specific assay so that assay-specific forms of the GH-2000 score are available and the basis of the data generating process. Currently, there are three IGF-I assays and two P-III-NP assays approved by WADA. The IGF-I assays are:

- Liquid chromatography-tandem mass spectrometry (LC-MS/MS)
- Immunotech A15729 IGF-I IRMA (Immunotech SAS, Marseille, France)
- Immunodiagnostic Systems iSYS IGF-I (Immunodiagnostics Systems Limited, Boldon, UK)

The P-III-NP assays are:

- UniQ$^{\text{IM}}$ P-III-NP RIA (Orion Diagnostica, Espoo, Finland)
- Siemens ADVIA Centaur P-III-NP (Siemens Healthcare Laboratory Diagnostics, Camberley, UK).

For more details and background on these assays, see Holt et al.[5] As any GH-2000 score requires a pair of IGF-I assay and P-III-NP assay there are six GH-2000 scores possible to construct. Depending on the available technology, laboratories choose the appropriate GH-2000 score for evaluating their samples and decision limits have been developed for all six GH-2000 scores in Holt et al.[5].

Intra-subject and inter-subject variability are important issues for the performance of the diagnostic test and have been considered carefully during its development. The intra-subject variability can be divided into biological and analytical variability, both of which have been studied. We have previously investigated the intra-subject biological variability of the GH-2000 scores and its components. The intra-individual variability for IGF-I ranged between 14-16% while the variability for P-III-P was 7-18%[6]. The overall mean intra-individual variability of the GH-2000 score was less than 0.6 units. Furthermore within-assay and between-assay variability has been determined by the WADA-accredited Drug Control Centre at Kings College London. In order to minimise assay variability between laboratories, WADA also assesses laboratory performance by the use of quality control samples before approving them to undertake this test.

Currently the test is designed to be used on a one-off basis and for this, within-subject variability is less relevant. However, the anti-doping authorities are developing a biological passport whereby each individual acts as his or her own control. The biological passport currently has haematological and steroidal modules but it is likely that this will be expanded to include an endocrine module[7]. The premise of the test is based on the variability between athletes. We have studied the mean and SD of a large population of elite athletes in order to determine when the GH-2000 is not biologically plausible without doping.

Note that the GH-2000 score consist of a linear discriminant function containing the two IGF-I and P-III-NP assay measurements and an age correction term. The age correction is required because GH secretion and markers of its action rise during childhood and reach a peak in early adulthood before declining again.[12] Without an adjustment for age, younger athletes are placed at a disadvantage. The inverse term for age is designed to adjust for age so that the score becomes independent of age. This is important in order to make the test applicable to athletes of all ages. Again details are available including an improved age-adjustment in Böhning et al.[13].

In the following, we assume that a GH-2000 score from an elite athlete without GH misuse is available and denote it by X. We assume further, and this is supported by substantial empirical evidence, that X of an athlete without GH misuse is normally distributed. Let the mean of X be denoted by μ and the variance of X by σ^2. Due to a requirement of WADA[2] it is desired to have a false-positive rate (FPR) of 1 in 10,000 which leads to a critical value c with the property that the probability for X exceeding c is 1 in 10,000, that is,

$$P(X > c) = 1/10000. \tag{2}$$

See also Figure 1 for an illustration. We can write (2) as

$$P\left(\frac{X-\mu}{\sigma} > \frac{c-\mu}{\sigma}\right) = 1/10000, \tag{3}$$

which can further be written as

$$1 - \Phi((c-\mu)/\sigma)) = 1/10000, \tag{4}$$

where $\Phi(z)$ is cumulative distribution function of a standard normal distribution. As $\Phi(z) = 0.9999$ is solved by $z = 3.72$, we find

$$(c-\mu)/\sigma = \Phi^{-1}(0.9999) = 3.72, \tag{5}$$

or, equivalently

$$c = \mu + 3.72\sigma. \tag{6}$$

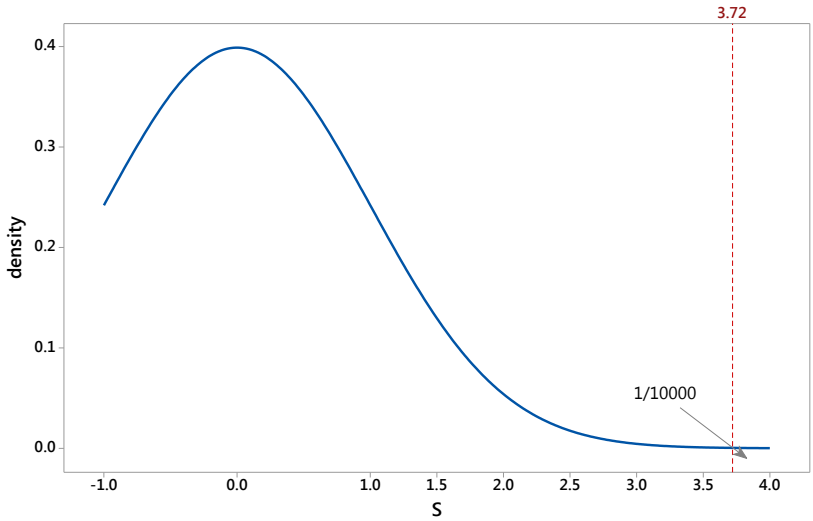

Figure 1. Normally distributed GH-2000 score with critical value for achieving a 1 in 10,000 false-positive rate

Clearly, had μ and σ been known, $c = \mu + 3.72\sigma$ would be the decision limit for us as it has a 1 in 10,000 FPR. However, μ and σ are unknown and need to be estimated by their sample estimates $\bar{x}$ and s, the sample mean and sample standard deviation, respectively. For that purpose, it is assumed that a sample of GH-2000 scores from elite athletes without GH misuse of size n is available. Furthermore, each elite athletes sample was only analysed once so that n refers to number of athletes.

As we need to replace μ and σ^2 by their respective sample estimates we only achieve an estimated critical value $\hat{c}$, namely

$$\hat{c} = \bar{x} + 3.72s. \tag{7}$$

As $\hat{c}$ is a random quantity, there is no guarantee that $\bar{x} + 3.72s \geq \mu + 3.72\sigma$ (in order to ensure at most 1 in $10,000$ FPR). Hence it is required to incorporate this uncertainty into the construction of the decision limit. This can be accomplished by using an additional uncertainty term u such that

$$P(\bar{X} + 3.72S + u \geq \mu + 3.72\sigma) = 0.95, \tag{8}$$

that is, we aim to construct an upper 95% confidence interval for the critical value c associated with the exact 1 in 10,000 FPR. In other words, we would like to construct an upper value, so that the true critical value associate with a FPR of 1 in 10,000 is equal or below this upper value with a pre-specified confidence level 0.95. Now, (8) is equivalent with

$$P\left(\frac{\bar{X}+kS-\mu-k\sigma}{\sqrt{Var(\bar{X}+kS)}} \geq -u/\sqrt{Var(\bar{X}+kS)}\right) = 0.95, \tag{9}$$

where we set $k = 3.72$. Assuming that $\bar{X}+kS$ is (approximately) normal we can consider (9) and solve it for u, namely $-u/\sqrt{Var(\bar{X}+kS)} = -1.65$, or

$$u = 1.65\sqrt{Var(\bar{X}+kS)}. \tag{10}$$

What remains is to find the variance of $\bar{X}+kS$. In the appendix, we recall the arguments leading to the following approximation of $Var(\bar{X}+kS)$, which can be found under normality to be

$$Var(\bar{X}+kS) \approx \frac{\sigma^2}{n}(1+k^2/2), \tag{11}$$

and $\frac{\sigma^2}{n}(1+k^2/2)$, the right-hand side of (11), can be readily estimated by $\frac{s^2}{n}(1+k^2/2)$. Hence we are able to provide a more concrete form of (10), namely

$$u = 1.65s\sqrt{(1+k^2/2)/n} \tag{12}$$

leading to the final decision limit

$$\bar{x}+ks+\underbrace{1.65s\sqrt{(1+k^2/2)/n}}_{\text{uncertainty adjustment}} = \bar{x}+\underbrace{[k+1.65\sqrt{(1+k^2/2)/n}]}_{\text{general multiplier}}\,s. \tag{13}$$

Clearly, with increasing sample size n, the uncertainty adjustment will loose its importance as it is inversely related to $\sqrt{n}$, and the decision limit simply becomes $\bar{x}+ks$ as n approaches infinity.

3 An alternative and exact approach

To rephrase the problem, we need h such that

$$P(\bar{X}+hS \geq \mu+k\sigma) = 0.95, \tag{14}$$

where $X_1, \cdots, X_n$ are independent and identically distributed, each having a normal distribution $X_i \sim N(\mu, \sigma^2)$ for $i = 1, \cdots, n$. It follows that $\bar{X} \sim N(\mu, \sigma^2/n)$ and $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$, the latter denoting a Chi-Square distribution with $n-1$ degrees of freedom. From the expression in (14), a reader who is familiar with the statistical methodology of tolerance sets (see for example Hahn and Meeker[17] or Krishnamoorthy and Mathew[16]) will easily realize that what we want is the 0.9999-content 95%-confidence upper tolerance limit for a normal distribution, which is also a special case of inference about quantiles of a normal distribution[15].

Next, for the convenience of the reader, we give a simple derivation of this upper tolerance limit. Logically equivalent transformations of the statement $\bar{X} + hS \geq \mu + k\sigma$ in (14) lead to

$$\bar{X} - \mu \geq k\sigma - hS \tag{15}$$

and

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq \frac{k\sigma - hS}{\sigma/\sqrt{n}}, \tag{16}$$

or

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq \sqrt{n}k - \frac{hS\sqrt{n}}{\sigma}, \tag{17}$$

and finally

$$Z - \sqrt{n}k \geq -\frac{hS\sqrt{n}}{\sigma}. \tag{18}$$

Note that Z is a standard normal random variable: $Z \sim N(0,1)$. Hence (18) becomes equivalent to

$$T = \frac{Z - \sqrt{n}k}{S/\sigma} \geq -h\sqrt{n}. \tag{19}$$

From the definition of the non-central t-distribution given in the appendix, it is clear that $T = \frac{Z-\sqrt{n}k}{S/\sigma}$ has a non-central t-distribution with $(n-1)$ df and parameter of non-centrality $-\sqrt{n}k$. Now it is easy to determine h to yield that $P(T \geq -h\sqrt{n}) = 0.95$ which is equivalent to $P(T < -h\sqrt{n}) = 0.05$. Hence we can find h as $h = -\Psi^{-1}_{n-1;\sqrt{n}k}(0.05)/\sqrt{n}$, namely the 5-th percentile of the non-central t-distribution with $(n-1)$ degrees of freedom and parameter of non-centrality $-\sqrt{n}k$. Here $\Psi_{df;\delta}(t)$ denotes the cumulative distribution function of the t-distribution with df degrees of freedom and parameter of non-centrality δ. This leads to the new form of decision limit as

$$\bar{x} + hs \tag{20}$$

where $h = -\Psi^{-1}_{n-1;\sqrt{n}k}(0.05)/\sqrt{n}$.

Table 1. The new multiplier $h = -\Psi^{-1}_{n-1;\sqrt{n}k}(0.05)/\sqrt{n}$ and the conventional multiplier $k + 1.65\sqrt{(1 + k^2/2)/n}$ in dependence of the sample size n

n	$k + 1.65\sqrt{(1 + k^2/2)/n}$	$-\Psi^{-1}_{n-1;\sqrt{n}k}(0.05)/\sqrt{n}$
5	5.7965	8.9683
10	5.1883	6.2205
20	4.7583	5.1681
50	4.3767	4.5143
100	4.1843	4.2476
200	4.0483	4.0781
500	3.9277	3.9388
1,000	3.8668	3.8722
2,000	3.8238	3.8263
100,000	3.7347	3.7347
∞	3.7190	3.7190

It is interesting to compare the decision limit in (20) and the decision limit used currently:

$$\bar{x} + [k + 1.65\sqrt{(1 + k^2/2)/n}]s. \tag{21}$$

Table 1 presents some comparative results. It can be seen that there are considerable differences between the two for small sample sizes but these become small for $n = 100$ and above. The differences become negligible for $n = 1,000$ and above.

4 True false positive rate

If the true population values μ and σ^2 would be known, a decision limit of the form $\mu + 3.72\sigma$ would provide an exact false positive rate of 1 in $10,000$. However, μ and σ^2 have to be estimated from a sample leading to statistical error. To adjust for this error an upper 95% confidence limit is chosen as the decision limit. In fact, the decision limit has been developed as $\bar{X} + hS$ where h is either given in (20) or in (21). If X represents a new determination of the GH-2000 score, then the *true false positive rate* (TFPR) is defined as

$$P(X > \bar{X} + hS),$$

where $\bar{X}$ and S are sample mean and sample standard deviation of a sample of GH-2000 score measurements $X_1, \cdots, X_n$ for n athletes. Note that X is not part of the sample but assumed to arise

Table 2. TFPR per 10,000 for $h = -\Psi^{-1}_{n-1;\sqrt{n}k}(0.05)/\sqrt{n}$ and the conventional multiplier $h = k + 1.65\sqrt{(1+k^2/2)/n}$ in dependence of the sample size n

n	$h = k + 1.65\sqrt{(1+k^2/2)/n}$	$h = -\Psi^{-1}_{n-1;\sqrt{n}k}(0.05)/\sqrt{n}$
5	30.6115	6.0624
10	3.9735	1.1023
50	0.3632	0.2317
100	0.3348	0.2645
1000	0.5915	0.5790
2000	0.6798	0.6730
100,000	0.9403	0.9403
∞	1.0000	1.0000

from the same population the sample has come from. Note that we have the equivalence

$$P(X > \bar{X} + hS) = P([X - \bar{X}]/S > h),$$

so that interest is in $[X - \bar{X}]/S$. Now, $(X - \bar{X})$ is normal with variance $\sigma^2(1 + \frac{1}{n}) = \sigma^2\frac{n+1}{n}$. Hence $Z = \frac{(X-\bar{X})}{\sigma}\sqrt{\frac{n}{n+1}} \sim N(0,1)$ is a standard normal random variable. In addition, $(n-1)S^2/\sigma^2 \sim \chi^2_{(n-1)}$ is distributed as χ^2 with $n-1$ degrees of freedom. This leads to the fact that

$$\frac{Z}{S/\sigma} = \frac{(X - \bar{X})}{S}\sqrt{\frac{n}{n+1}} \sim t_{(n-1)}, \tag{22}$$

has a t-distribution with $(n-1)$ degrees of freedom. It follows that

$$P(\frac{X - \bar{X}}{S} > h) = P(\frac{X - \bar{X}}{S}\sqrt{\frac{n}{n+1}} > \sqrt{\frac{n}{n+1}}h) = 1 - \Psi_{n-1}\left(\sqrt{\frac{n}{n+1}}h\right) \tag{23}$$

where $\Psi_\nu(.)$ is the cumulative distribution function of a t-distribution with ν degrees of freedom. Hence we have an easy formula for calculating the *true false positive rate.*

In Table 2 we have calculated the TFPR for selected values of n. We see that the exact method (20) meets the 1 in 10,000 target even for small sample sizes whereas the conventional method (13) needs slightly larger sample sizes. It is pointed out here that, for both methods, convergence to the target value of 1 in 10,000 is *from below* which emphasizes that the uncertainty term is chosen appropriately.

5 Discussion

We have developed a new and exact form of decision limits to cope with uncertainty in the decision limit estimate based on the GH-2000 score for growth hormone misuse. The decision limits currently in use, which were developed using the large sample method, are based on sample sizes of at least 1000. Hence the new multiplier, developed on the basis of the non-central t-distribution, is not substantially different for large sample sizes and there is no need to modify the current decision limits as implemented in the WADA guidelines as these are based on large sample sizes. However, the new approach is still valuable. The conventional approach uses approximations at two places:

- it is assumed that $\bar{X} + kS$ is normally distributed. This will only hold approximately and the approximation will be not good for small sample size. Of course, using the central limit theorem here, the approximation is well defendable for larger sample sizes. However, there is always debate on what establishes a large sample size.
- in computing the variance of S we are using a further approximation by means of applying the δ-method. This might be reasonable or not. Recent simulation work shows that for sample sizes of 1,000 and above the $\delta-$method based approximation discussed in the appendix is quite reasonable.[14]

The major benefit of the new approach is making these two approximations unnecessary. It is clear that these decision limits more defendable, for example in court cases, as they are exact for any sample size n, large or small. Clearly, a large sample size (such as 1000 or more) is preferable to a small sample size. However, it needs to be seen that the conventional method (13) still uses a normality assumption for its development (as does the new method (20)), but the new method avoids the approximations mentioned above.

It is interesting to note though that for small sample sizes the new approach is more suitable. Hence this might be the way to proceed when new assays come into use replacing older ones and new decision limits have to be calculated.

Another issue is the choice of the upper 95% confidence limit as coping with the statistical uncertainty. This was suggested in Erotokritou-Mulligan et al.,[11] presumingly for no other reason than that this is the standard statistical significance level. In addition, this proposal was accepted by WADA and has become part of their guidelines[2]. Clearly, increasing the confidence level would further enlarge the decision limit, but this would also lead to an increase in the false negative rate and in the investigation of the TFPR in section 4 clearly shows that this is not necessary as the current confidence level ensures convergence to the true false positive rate *from below*.

It is clear that a larger decision limit will increase the false negative rate (FNR). Hence the conventional method (13) might have a smaller FNR as the associated decision limit is smaller than the exact method

(20), but this is purely because the conventional method is approximate and has a larger FPR than the exact method which guarantees the 1/10,000 FPR for any reasonable sample size. The FNR has been investigated previously[10,19]; here we are not able to make any statements on the FNR as this would require GH-2000 scores from confirmed positive athletes.

Appendix: finding the variance of the sample standard deviation S

It is known that, under normality, $Var(s^2) = 2\sigma^4/(n-1)$. This follows from the fact that

$$(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{(n-1)},$$

a chi-square distribution with $n-1$ degrees of freedom. Hence

$$Var[(n-1)\frac{S^2}{\sigma^2}] = 2(n-1)$$

and the result follows.

Recall that the $\delta-$ method implies that if $T(X)$ is a differentiable transformation of X then

$$Var(T(X)) \approx T'(E(X))^2 Var(X).$$

This result is applied with $T(S^2) = \sqrt{S^2}$ to give:

$$Var(S) = Var(\sqrt{s^2}) \approx Var(S^2)\frac{1}{4}(\frac{1}{\sqrt{\sigma^2}})^2 = \frac{1}{2}\sigma^2/(n-1),$$

which is the result we have used in Section 2.

Appendix: non-central t-distribution

Recall that if $Z \sim N(0,1)$ and independent $V \sim \chi^2_\nu$ then

$$T = \frac{Z+\mu}{\sqrt{V/\nu}}$$

has a t-distribution with parameter of non-centrality μ and v degrees of freedom. In our case, we choose $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ and $V = (n-1)S^2/\sigma^2$ as well as $\nu = (n-1)$ and $\mu = -\sqrt{n}k$ so that $T = \frac{z+\mu}{\sqrt{V/\nu}} = \frac{z-\sqrt{n}k}{s/\sigma}$ which is used in section 3. For more details on the noncentral t-distribution see Johnson et al.[18].

References

1. Holt, RI. Is human growth hormone an ergogenic aid? Drug Test Anal 2009 Sep 2009. doi: 10.1002/dta.58.
2. World Antidoping Agency. World Anti-Doping Program Guidelines for hGH Isoform Differential Immunoassays for anti-doping analyses. https://wada-main-prod.s3.amazonaws.com/resources/files/WADA-Guidelines-for-hGH-Differential-Immunoassays-v2.1-2014-EN.pdf . 2014.
3. WADA The World Anti-Doping Code International Standard: Prohibited List 2016, https://wada-main-prod.s3.amazonaws.com/resources/files/wada-2016-prohibited-list-en.pdf
4. Bidlingmaier M, Wu Z, Strasburger CJ. Test method: GH. Baillieres Best Pract Res Clin Endocrinol Metab. 2000 Mar;14(1):99-109.
5. Holt RI, Böhning W, Guha N, Bartlett C, Cowan, DA, Giraud S, Bassett EE, Sönksen PH, Böhning D. The development of decision limits for the GH-2000 detection methodology using additional insulin-like growth factor-I and amino-terminal pro-peptide of type III collagen assays. Drug Test Anal 2015 Jan 21. doi: 10.1002/dta.1772.
6. Erotokritou-Mulligan I, Bassett EE, Cowan DA, Bartlett C, Milward P, Sartorio A, Sönksen PH, Holt RIG. The use of growth hormone (GH)-dependent markers in the detection of GH abuse in sport: Physiological intra-individual variation of IGF-I, type 3 pro-collagen (P-III-P) and the GH-2000 detection score. Clinical Endocrinology 2010; 72: 520-526.
7. Sottas PE, Vernec A. Current implementation and future of the Athlete Biological Passport. Bioanalysis 2012; 4(13): 1645–1652. doi: 10.4155/bio.12.130.
8. Longobardi S, Keay N, Ehrnborg C, Cittadini A, Rosen T, Dall R, et al. Growth hormone (GH) effects on bone and collagen turnover in healthy adults and its potential as a marker of GH abuse in sports: a double blind, placebo-controlled study. The GH-2000 Study Group. J Clin Endocrinol Metab 2000 Apr;85(4):1505-12.
9. Dall R, Longobardi S, Ehrnborg C, Keay N, Rosen T, Jorgensen JO, et al. The effect of four weeks of supraphysiological growth hormone administration on the insulin-like growth factor axis in women and men. GH-2000 Study Group. J Clin Endocrinol Metab 2000 Nov;85(11):4193-200.
10. Powrie JK, Bassett EE, Rosen T, Jorgensen JO, Napoli R, Sacca L, Christiansen JS, Bengtsson BA, Sönksen PH. Detection of growth hormone abuse in sport. Growth Horm IGF Res 2007 17: 220-226.
11. Erotokritou-Mulligan I, Guha N, Stow M, Bassett EE, Bartlett C, Cowan DA, Sönksen PH, Holt RIG. The development of decision limits for the implementation of the GH-2000 detection methodology using current commercial insulin-like growth factor-I and amino-terminal pro-peptide of type III collagen assays . Growth Hormone & IGF Research 2012 doi:10.1016/j.ghir.2011.12.005

12. Toogood AA. Growth hormone (GH) status and body composition in normal ageing and in elderly adults with GH deficiency. Horm Res. 2003;60(Suppl 1):105-11.
13. Böhning D, Böhning W, Guha N, Cowan DA, Sönksen PH, Holt RIG. Statistical methodology for age-adjustment of the GH-2000 score detecting growth hormone misuse. BMC Medical Research Methodology 2016; 16:147 DOI: 10.1186/s12874-016-0246-8.
14. Kalevras, M. Statistical Methodology for Growth Hormone Misuse Detection. MSc thesis, School of Mathematical Sciences, University of Southampton, 2016.
15. Liu W, Bretz F, Hayter A, Glimm E. Simultaneous inference for several quantiles of a normal population with applications. Biometrical Journal 2013; 55(3): 360–369.
16. Krishnamoorthy K, and Mathew, T. Statistical Tolerance Regions – Theory, Applications, and Computation. New York: Wiley, 2009.
17. Hahn G, and Meeker W Q. Statistical Intervals: A Guide to Practitioners. New York: Wiley, 1991.
18. Johnson N L, Kotz S, Balakrishnan N. Continuous Univariate Distributions Volume 2, 2nd edition. New York: Wiley, 1996.
19. Holt RIG, Erotokritou-Mulligan I, McHugh C, Bassett EE, Bartlett C, Fitvan A, Bacon JL, Cowan DA, Sönksen PH. The GH-2004 project: the response of IGF1 and type III pro-collagen to the administration of exogenous GH in non-Caucasian amateur athletes. Eur J Endocrinol. 2010; 163(1): 45–54. doi: 10.1530/EJE-09-0978.