Statistical Methodology for Age-Adjustment of the GH2000 Score Detecting Growth Hormone Misuse

S3RI Methodology Working Paper, February 2016

Dankmar **Böhning**[1], Walailuck **Böhning**[2], Nishan **Guha**[2,3] David A**. Cowan**[4], Peter H. **Sönksen**[2] and Richard I.G**. Holt** [2]

1) Southampton Statistical Sciences Research Institute, University of Southampton, UK

2) Human Development and Health Academic Unit, University of Southampton Faculty of Medicine, Southampton, UK

3) Nuffield Division of Clinical Laboratory Sciences, University of Oxford, UK

4) Department of Pharmacy and Forensic Science, Drug Control Centre, King's College London, UK

Statistical Methodology for Age-Adjustment of the GH2000 Score Detecting Growth Hormone Misuse

Dankmar **Böhning**[1], Walailuck **Böhning**[2], Nishan **Guha**[2,3] David A**. Cowan**[4], Peter H. **Sönksen**[2] and Richard I.G**. Holt** [2]

1) Southampton Statistical Sciences Research Institute, University of Southampton, UK

2) Human Development and Health Academic Unit, University of Southampton Faculty of Medicine, Southampton, UK

3) Nuffield Division of Clinical Laboratory Sciences, University of Oxford, UK

4) Department of Pharmacy and Forensic Science, Drug Control Centre, King's College London, UK

Abstract

Background. The GH2000 score has been developed as a powerful and unique technique for the detection of growth hormone misuse by sportsmen and women. With the collection and establishment of an increasingly large data base it has become apparent that the score shows a positive age effect in the male athlete population, which could potentially place older male athletes at a disadvantage.

Methods. We have used results from residual analysis of the general linear model to show that the residual of the GH2000 score when regressed on the mean-age centered age is the right way to proceed to correct this bias. As six GH2000 scores are possible depending on the assays used for determining IGF-I and P-III-NP, methodology had to be explored for including six different age effects into a unique residual. Meta-analytic techniques have been utilized to find a summary age effect.

Results. This form of age-adjusted GH2000 score, a form of residual, has similar mean and variance as the original GH2000 score and, hence, the developed decision limits show negligible change when compared to the decision limits based on the original score. We also show that any further scale-transformation will not change the adjusted score any more. Hence the suggested adjustment is optimal for the given data. The summary age effect is homogeneous across the six scores, and so the generic adjustment of the GH2000 score formula is justified.

Conclusions. A final revised GH2000 score formula is provided which is independent of the age of the athlete under consideration.

Key words: GH2000 score, adjusting for age effects, meta-analysis of scores, centering and norming of scores

Introduction and background

Growth hormone is a powerful anabolic agent of considerable therapeutic value but also misused in sport. In order to preserve the fairness of competition 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 Strasburger, Wu and Bidlingmaier and the GH2000 biomarker test developed by the GH2000 and GH-2004 projects (Holt *et al.* 2015). 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). 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 GH2000 score) on which basis the compliance of the sample's analytical result is determined.

The initial development of the GH2000 score was reported by Powrie *et al.* (2007) but this score was based on immunoassays that are no longer commercially available. Although the original discriminant function has remained unchanged, the decision limits have been updated as further experience was accumulated and new assays became available (Erotokritou-Mulligan *et al.* 2012 and Holt *et al.* 2015). Currently, there are three IGF-I assays and two P-III-NP assays approved by WADA.

The IGF-I assays used in this study were:

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

The P-III-NP assays used in this analysis were:

- UniQ™ 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.* (2015).

As these assays do not give identical results, different GH2000 scores are obtained with each of the combinations and this means that the decision limits are slightly different, depending on the assay pair used.

The score itself involves measurement of IGF-I and P-III-NP in a linear way and a term that involves age inversely

$$\text{GH2000 score} = \beta_0 + \beta_1 \log\left(\text{IGF-I}\right) + \beta_2 \log\left(\text{P-III-NP}\right) + \beta_3 / \text{age} \qquad (1)$$

where the coefficients $\beta_0, \beta_1, \beta_2, \beta_3$ have different values for male and female athletes.

The inverse term for age in (1) 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. Recent analysis of a combined database of elite athletes (Holt *et al.* 2015) provides evidence that the score is independent of age for the female population whereas it shows a linear dependence for male athletes. The combined database contains blood samples of athletes collecting at various sporting events including the 2011 International Association of Athletics Federations (IAAF) World Athletics Championships in Daegu, South Korea, in the following abbreviated as the Daegu-sample.

Figure 1 shows the scores and their relationship to age in 597 male athletes competing in Daegu. There are 6 scores as there are 3 assays for IGF-I (LC-MS/MS, Immunotech, IDS) and 2 for P-IIIN-P (Siemens-Centaur, Orion). It is clear from Figure 1 that in all GH2000-scores there is a positive age dependency as all linear regression lines show a significant age-effect. There is no age effect on the GH2000 scores for the female population of the Daegu-sample (data not shown).

The purpose of this paper is to suggest and discuss statistical methodology for adjusting the existing score for the undesirable age-effect.

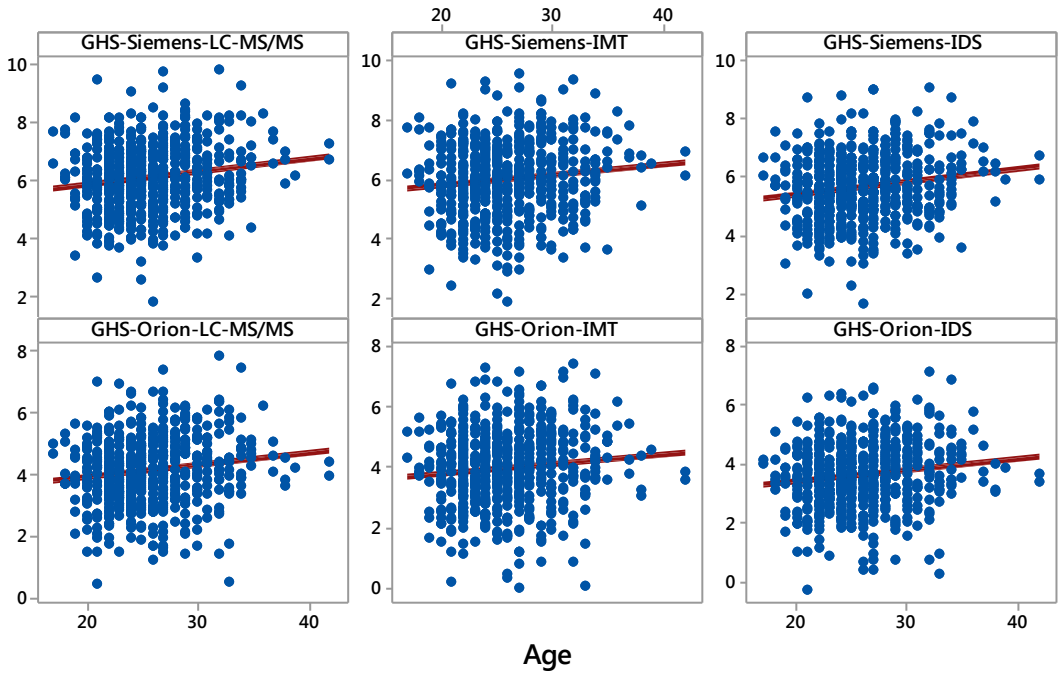

Figure 1: Scatterplots with regression lines for the six GH-scores available of all *male* athletes in the Daegu-sample in the order of their appearance: Siemens-LC-MS/MS (top-left), Siemens-Immunotech (top-middle), Siemens-IDS (top-right), Orion-LC-MS/MS (bottom-left), Orion-Immunotech (bottom-middle), Orion-IDS (bottom-right)

Methods

The basics of adjustment

Consider a response Y (in our case the GH2000 score) and an effect x (in our case the age of an athlete). Suppose that the response Y is related to x by a linear regression model

$$E(Y) = \alpha + \beta x$$

Then, the least-squares estimate of β is given by

$$\beta = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where the pairs (Y_i, x_i) represent the n sample values of Y and x. On this basis we are able to construct a response Y^* adjusting for x

$$Y^* = Y - \hat{\beta} x \ .$$

The adjusted response Y^* is independent of x as the following analysis shows. Tthis can be found in most books on regression but it is mentioned here for completeness. Consider the least-squares-estimate of β^* in

$$E(Y^*) = \alpha^* + \beta^* x .$$

This is given as the least-squares estimate of β^*

$$\beta^* = \frac{\sum_{i=1}^{n}(Y_i^* - \overline{Y^*})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(Y_i - \beta x_i - (\bar{Y} - \beta \bar{x}))(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(x_i - \bar{x}) - \beta \sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} - \beta \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = 0 .$$

Hence Y^* is independent of x. A more general result is provided in Appendix A.

Next, we suggest considering an adjustment of the form

$$Y^* = Y - \beta(x - \bar{x}) .$$

The benefit of this adjustment lies in the fact that the adjusted score Y^* remains on the *same level* as the original score Y as

$$\bar{Y}^* = \bar{Y} - \beta(\bar{x} - \bar{x}) = \bar{Y}.$$

The process of considering $x - \bar{x}$ is called *centering*. Sometimes also *norming* is considered in addition to centering which is $(x - \bar{x}) / sd(x)$ where $sd(x) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x - \bar{x})^2}$. We are not considering norming here as this will not lead to any further adjustment. To see this, considered any scale transformation ax of x. The original model $E(Y) = \alpha + \beta x$ becomes now $E(Y) = \alpha^* + \beta^* x^*$, where $x^* = ax$. Then, least squares estimates can be found as

$$\beta^* = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(x_i^* - \bar{x^*})}{\sum_{i=1}^{n}(x_i^* - \bar{x^*})^2} = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(x_i - \bar{x})a}{\sum_{i=1}^{n}(x_i - \bar{x})^2 a^2} = \frac{1}{a}\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{1}{a}\beta$$

Hence the adjusted response

$$Y^* = Y - \beta^* x^* = Y - (\frac{1}{a}\beta)ax = Y - \beta x$$

is indeed identical to the original adjustment $Y - \beta x$ and does not lead to anything new. A more general result is provided in Appendix B. Hence we stay with the adjustment

$$Y^* = Y - \beta(x - \bar{x})$$

as the final form of adjustment.

Adjusting the GH2000 score

To adjust the GH2000 score, we consider the regression of the GH2000 score on age. Table 1 shows 6 age-effects for the 6 GH2000 scores (as there are 2 assays for measuring PIIINP and 3 assays for measuring IGF-I).

Table 1: Estimated β-coefficients of the age-effects for the six GH2000 scores and their associated standard errors

GH2000 score		β age	S.E.
P-III-NP assay	IGF-I assay		
Siemens	LC-MS/MS	0.0418	0.0082
Siemens	Immunotech	0.0261	0.0086
Siemens	IDS	0.0359	0.0070
Orion	LC-MS/MS	0.0363	0.0077
Orion	Immunotech	0.0202	0.0085
Orion	IDS	0.0318	0.0077

For simplicity and ease of use by the anti-doping laboratories, it is important that we do not create an age adjustment for each assay pairing. Thus we need to include the age adjustment within the generic GH2000 score (independent of the specific assay pairing used). To accomplish this task we have applied ideas from meta-analysis. We consider each GH2000 score using a specific assay combination as a realisation from multiple possible assay combinations.

This is similar to a meta-analysis approach in which studies aiming to estimate a certain effect are considered as realisation from a universe of possible studies.

Hence we use

$$\overline{\beta} = \sum_{i=1}^{k} w_i \beta_i \Big/ \sum_{i=1}^{k} w_i$$

where k =6 is the number of different assay combinations used and $\hat{\beta}_i$ is the estimated age effect, and w_i is the inverse of the estimated variance (the squared values in column 3 of table 1). Hence $\overline{\beta}$ is an average of the estimated effect.

Results

In our case, we find $\overline{\beta}$ = 0.032. Figure 2 shows this analysis graphically. As all assay-specific age effects are similar in their standard error, all weights are similar. More details on the meta-analysis approach are given in Appendix C.

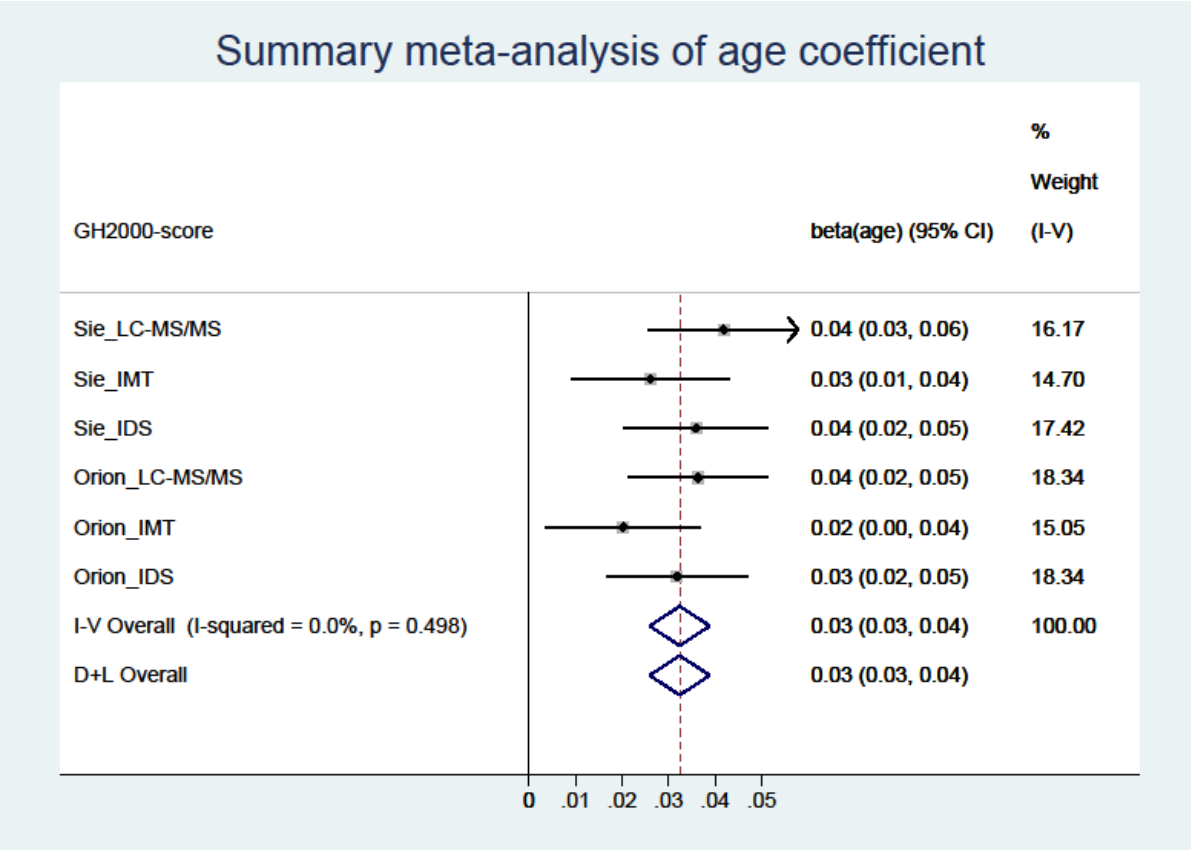

Figure 2: Meta-analytic results for the six age-effects of the GH2000 scores on age (I-V stands for overall inversely weighted and provides the summary estimate of the age-affect); more details are given in the appendix

From the meta-analysis, we achieve the formula for the male athletes:

GH2000 score-adj = GH2000 score – 0.032 (age - 25.09)

AS the mean age for male athletes is 25.09 years and the GH2000 is calculated as:

GH2000 score = - 6.586 + 2.905 log(P-III-NP) + 2.100 log(IGF-I) – 101.737/age

the *adjusted score* formula becomes:

GH2000 score-adj = - 5.783 +2.905 log(P-III-NP) + 2.100 log(IGF-I)

– 101.737/age – 0.032 age. (2)

Figure 3 shows a scatterplot of the six *age-adjusted* GH2000-scores. It clearly shows that the age-effect is removed as it is expected from the above theory.

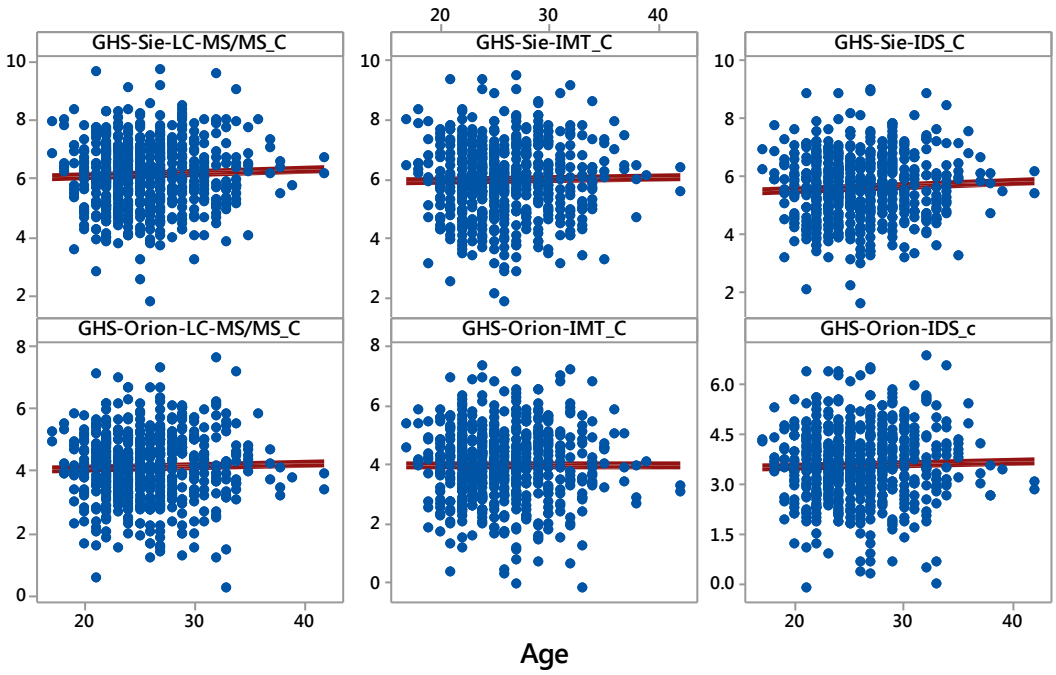

Figure 3: Scatterplots with regression lines for the six *age-adjusted* GH-scores of all *male* athletes in the Daegu-sample in the order of their appearance: Siemens-LC-MS/MS (top-left), Siemens-Immunotech (top-middle), Siemens-IDS (top-right), Orion-LC-MS/MS (bottom-left), Orion-Immunotech (bottom-middle), Orion-IDS (bottom-right)

Effect on the current WADA decision limits

Although this adjustment will lead to changes in the individual GH2000 score of an athlete, it has negligible effect on the decision limits. Following Holt *et al.* (2015) these are constructed using the 1 in 10,000 false positive rate as

$$\bar{y} + 3.72s + u$$

where $\bar{y}$ and s are mean and standard deviation of the respective GH2000 score. u is a sample uncertainly term defined as

$$u = \sqrt{\frac{s^2}{n}(1 + \frac{3.72^2}{n})}$$

where n is the sample size. Table 2 shows the details, in particular, a comparison between GH2000 scores with and without adjustment.

Table 2: Descriptive statistics including decision limits for the 6 unadjusted and adjusted GH2000 scores

Assay pair			n	mean	s	Mean+3.72*s	U	DL
P-III-NP	IGF-I							
Siemens	LC-MS	Unadjusted	947	6.5393	1.2412	11.1566	0.1872	11.34
		Adjusted	947	6.5382	1.2424	11.1599	0.1874	11.35
Siemens	Immunotech	Unadjusted	971	6.4292	1.3189	11.3355	0.1965	11.53
		Adjusted	971	6.4287	1.3284	11.3703	0.1979	11.57
Siemens	IDS	Unadjusted	970	5.9935	1.1925	10.4296	0.1777	10.61
		Adjusted	970	5.9931	1.1954	10.4400	0.1782	10.62
Orion	LC-MS	Unadjusted	966	4.7062	1.2902	9.5057	0.1927	9.70
		Adjusted	966	4.7047	1.2976	9.5318	0.1938	9.73
Orion	Immunotech	Unadjusted	999	4.5984	1.3925	9.7785	0.2045	9.98
		Adjusted	999	4.5984	1.4077	9.8350	0.2068	10.04
Orion	IDS	Unadjusted	992	4.1614	1.2522	8.8196	0.1846	9.00
		Adjusted	992	4.1611	1.2610	8.8520	0.1859	9.04

Conclusions

We are suggesting this adjustment for the male elite athlete population only, as the female population does not show age dependency.

The age-adjustment of the score is also beneficial with respect to the normality of the scores as the probability plot in figure 4 shows. All scores appear to be normal.

Another question relates to the appropriateness of the meta-analytic weighted average approach - are the age-effects for the six scores similar enough to be validly combined in a weighted average? This question is tackled with a heterogeneity analysis. The χ^2-test of homogeneity is

$$\chi^2 = \sum_{i=1}^{6} \frac{(\beta_i - \overline{\beta})^2}{\mathrm{var}(\hat{\beta}_i)} = 4.37$$

which has a non-significant p-value of 0.498 by 5 df. Hence the approach we have taken seems justified (details are given in the Appendix C).

Distribution of adjusted GH2000 scores

The construction of the decision limits for GH2000 biomarker methodology is dependent on a normal distribution of GH2000 scores among clean athlete. This was assessed using probability plotting and the Anderson-Darling test for normality which provided clear evidence that all six scores were normally distributed (Figure 4).

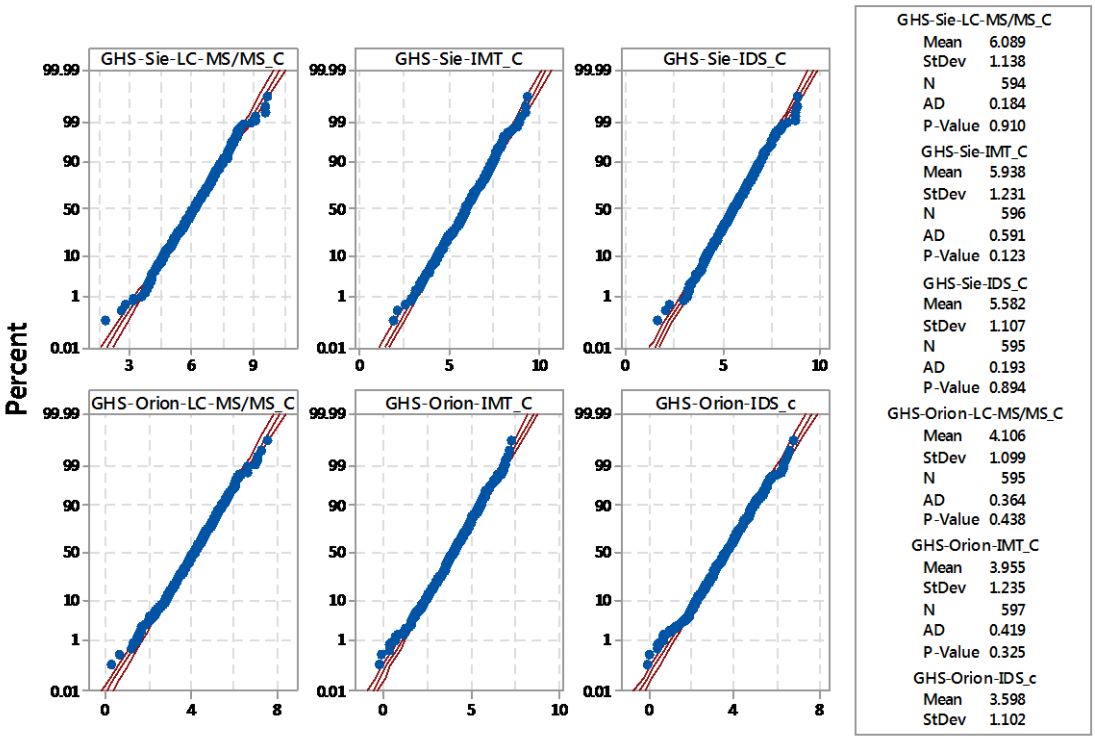

Figure 4: Probability plots for the six GH2000 scores adjusted for age

The GH2000 and GH2004 teams have previously published the rationale and background to the development of decision limits for the GH-2000 biomarker detection method (Powrie *et al.* 2007, Holt *et al.* 2015) .

It was always envisaged that a dynamic approach would be taken towards refining the decision limits as further data became available. Our recent investigations have shown that the age-adjustment in the male discriminant function, which was derived the original GH2000 cross-sectional elite athlete study, over-corrects for age in male athletes in our more recent cohorts. The effect of this over-correction is to place older male athletes at a slight disadvantage compared with their younger peers, for whom the sensitivity of the test is reduced. The original age correction for women remained valid in the later cohorts. We have used the most recent dataset, on which the current decision limits are based, to add a smaller further adjustment to the discriminant function to address this issue.

It is important to recognise that this age-correction was fitted to a specific cohort of elite athletes and only provides an estimate of the true age-correction. Subsequently we have found that the male age adjustment overcorrects for age, thus placing older male athletes at a small disadvantage. This has been addressed by this analysis. It is reassuring that the age correction remained valid for the female athletes in later cohorts

When undertaking this analysis, we used several principles to guide out work: 1) we wanted to ensure that the updated male discriminant function was unaffected by age in order to make the test equally fair and effective for athletes of all ages; 2) the change in age correction would have a minimal effect on the current decision limits; and 3) a single age

adjustment could be applied for all assay pairings. In order to minimise the effect on the current decision limits, we used a method that centered the data. By doing so the mean GH2000 scores were virtually unaffected. There was a trivial change to the SDs and consequently the decision limits, which are based on the mean and SD, were unchanged. The age adjustment varies slightly by assay pairing and in order to overcome this, we adapted meta-analytical methodology to derive a common age adjustment for all the combinations. There was no evidence of heterogeneity between the assay pairings and each contributed to the final adjustment equally, providing support for this approach.

In conclusion, we have created a small further age adjustment for male athletes to correct the age bias introduced with the original discriminant formula. This has no effect on the decision limits and should be easily introduced into anti-doping testing.

Appendix A: Independence of residuals from model covariates

Consider a general linear model $Y = X\beta + \varepsilon$ where $Y - X$ is a *n*-vector of responses, X is the design matrix containing the *n*-values of *p* covariates, ε is a *n*-vector of errors, and β is a *p*-vector of unknown parameters. Then, the least-squares estimate for β is given as

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

and the vector of residuals $Y^* = Y - X\hat{\beta}$. Regressing Y^* on X leads to the general linear model

$$Y^* = X\beta^* + \varepsilon^*$$

and the least-squares estimate of β^* is given as

$$\begin{aligned}
\hat{\beta}^* &= (X^T X)^{-1} X^T Y^* = (X^T X)^{-1} X^T (Y - X\hat{\beta}) \\
&= (X^T X)^{-1} X^T [Y - X(X^T X)^{-1} X^T Y] \\
&= (X^T X)^{-1} X^T Y - (X^T X)^{-1} X^T X (X^T X)^{-1} X^T Y \\
&= (X^T X)^{-1} X^T Y - (X^T X)^{-1} X^T Y = 0,
\end{aligned}$$

showing that the residuals are independent from all covariates included in the model. See also Sen and Srivastava (1990).

Appendix B: Invariance of the effect estimates with respect to scale transformations

Consider a general linear model $Y = X\beta + \varepsilon$ where Y is a *n*-vector of responses, X is the design matrix containing the *n*-values of *p* covariates, ε is a *n*-vector of errors, and β is a *p*-vector of unknown parameters. Now let A be an invertible $p \times p$ matrix and XA the associated scale-transformation of the design matrix. Then, the least-squares estimate of the transformed model $Y = XA\beta^* + \varepsilon^*$ is given as

$$\hat{\beta}^* = (A^T X^T XA)^{-1} A^T X^T Y = (X^T XA)^{-1} (A^T)^{-1} A^T X^T Y$$
$$= (X^T XA)^{-1} X^T Y = A^{-1} (X^T X)^{-1} X^T Y.$$

It follows that the residual with respect to the scale-transformed design matrix

$$Y - XA\hat{\beta}^* = Y - XAA^{-1} (X^T X)^{-1} X^T Y$$
$$= Y - X (X^T X)^{-1} X^T Y = Y - X\beta$$

is identical to the residual of the untransformed design matrix. See also Sen and Srivastava 1990. As a consequence norming (for example by standard deviations of covariates) of the covariates will not change the residuals.

Appendix C: Heterogeneity analysis

Here we give more details on the meta-analytic approach we have taken. Figure 5 shows the various elements involved in the meta-analysis.

The basic elements are the six GH2000 scores with their age-effects and weights according to the inverse variance (similar variance). The two bottom rows show the summary effect with and without heterogeneity. Both are virtually identical, as there is no heterogeneity (I^2 = 0, no variation due to heterogeneity). In case there is heterogeneity we would consider the DerSimonian-Laird approach which incorporates heterogeneity into the weighting scheme. In our case, both analyses lead to the same result. All analysis is based on the add-on package METAN of the statistical software STATA14 (Stata Corp. 2015).

Study	ES	[95% Conf.	Interval]	% Weight
Sie_LC-MS/MS	0.042	0.026	0.058	16.17
Sie_IMT	0.026	0.009	0.043	14.70
Sie_IDS	0.036	0.020	0.051	17.42
Orion_LC-MS/MS	0.036	0.021	0.051	18.34
Orion_IMT	0.020	0.004	0.037	15.05
Orion_IDS	0.032	0.017	0.047	18.34
I-V pooled ES	0.032	0.026	0.039	100.00
D+L pooled ES	0.032	0.026	0.039	100.00

Heterogeneity chi-squared = 4.37 (d.f. = 5) p = 0.498
I-squared (variation in ES attributable to heterogeneity) = 0.0%

Test of ES=0 : z= 9.82 p = 0.000

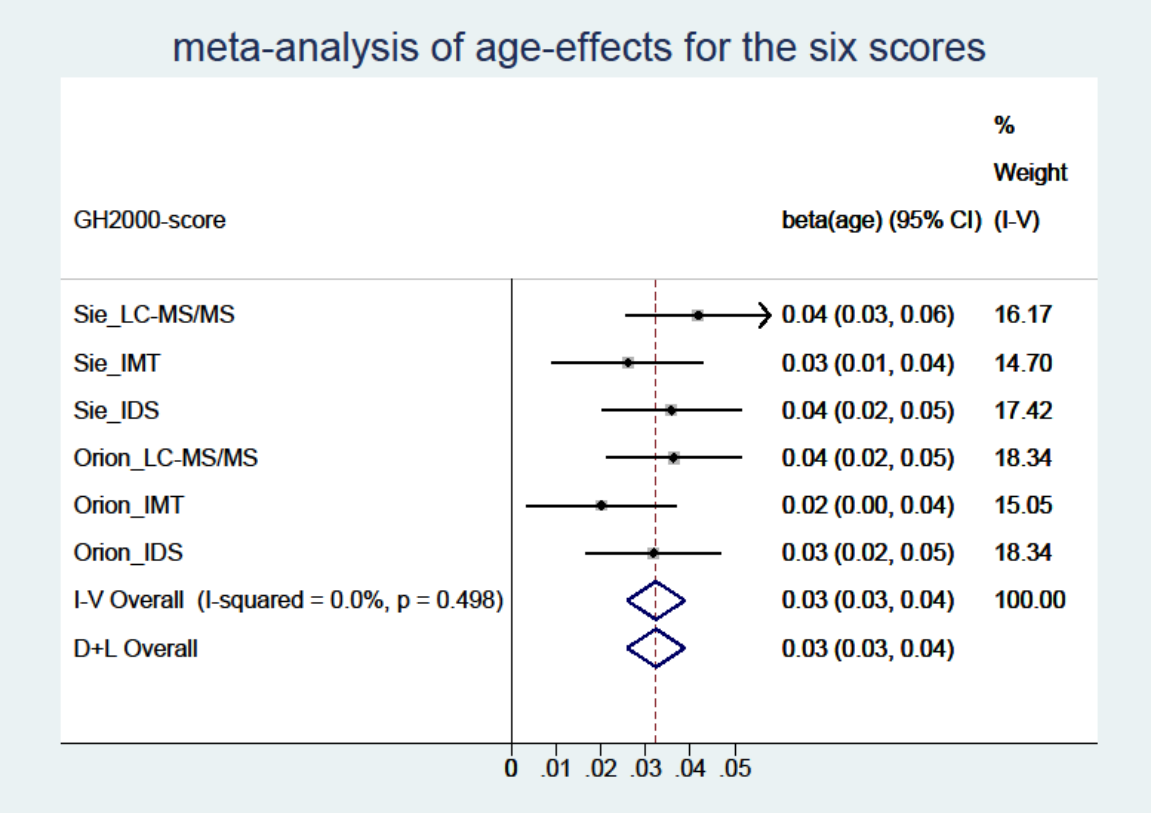

Figure 5: Meta-analytic displays including heterogeneity analysis based on METAN, an add-on package to the statistical software STATA14

Declarations

Acknowledgements

This work has received funding from WADA and the US Partnership for Clean Competition.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DB conceived the statistical theory of this study. WB carried out all computations and RH drafted the manuscript. NG, DC and PS critically reviewed and made substantial contributions to the manuscript. All authors commented on and approved the final manuscript.

References

Holt RI, Böhning W, Guha N, Bartlett C, Cowan, DA, Giraud S, Bassett EE, Sönsken 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.

Erotokritou-Mulligan I, Guha N, Stow M, Bassett EE, Bartlett C, Cowan DA, Sönksen PA, 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

Powrie JK, Bassett EE, Rosen T, Jorgensen JO, Napoli R, Sacca L, Christiansen JS, Bengtsson BA, Sonksen PH. Detection of growth hormone abuse in sport. *Growth Horm IGF Res* 2007 17: 220-226.

Sen A, Srivastava, M. *Regression Analysis: Theory, Methods and Applications.* Springer, Heidelberg 1990.

Stata Corp. . *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP 2015.