CLINICAL STUDY

The GH-2004 project: the response of IGF1 and type III pro-collagen to the administration of exogenous GH in non-Caucasian amateur athletes

Richard I G Holt, Ioulietta Erotokritou-Mulligan, Cathy McHugh, E Eryl Bassett 1, Christiaan Bartlett 2, Adam Fityan, Jenny L Bacon, David A Cowan 2 and Peter H Sönksen

Endocrinology and Metabolism, Developmental Origins of Health and Disease Division, The Institute of Developmental Sciences (IDS Building), MP887, Southampton General Hospital, University of Southampton School of Medicine, Tremoruna Road, Southampton SO16 6YD, UK, 1 School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury CT2 7NF, UK and 2 Department of Forensic Science and Drug Monitoring, King’s College London, Drug Control Centre, London SE1 9NH, UK

(Correspondence should be addressed to R I G Holt; Email: r.i.g.holt@southampton.ac.uk)

Abstract

Context: The GH-2000 team proposed a method based on IGF1 and type III pro-collagen (P-III-P) to detect exogenously administered GH. As previous studies involved predominantly white European athletes, it is important to assess whether the response of these markers to recombinant human GH (rhGH) differs with ethnicity.

Objective: To examine the response of serum IGF1 and P-III-P and GH-2000 score to rhGH in non-Caucasian amateur athletes.

Design: Double-blind placebo-controlled rhGH administration study.

Setting: Wellcome Trust Clinical Research Facility, Southampton General Hospital.

Subjects: The study included 31 male and 14 female amateur athletes of different ethnicities.

Intervention: The subjects were assigned to treatment with placebo or 0.1 IU/kg per day (low dose) or 0.2 IU/kg per day (high dose) rhGH for 28 days. Blood was collected weekly during treatment and on days 35, 42 and 84 during the washout period. Serum IGF1 and P-III-P were measured, and GH-2000 score was calculated.

Results: IGF1, P-III-P and GH-2000 score rose in response to both low- and high-dose GH in both men and women. When compared with the Caucasian volunteers of the previous GH-2000 study, mean baseline and placebo-treated P-III-P and GH-2000 score were lower in GH-2004 men and women. Post-GH, however, peak IGF1 or P-III-P did not differ between studies but the peak GH-2000 score was lower in GH-2004 men. There was no difference between studies in the maximal change in IGF1, P-III-P and GH-2000 score in response to GH in either gender.

Conclusions: These data do not support a significant ethnic effect on the peak or maximal response to rhGH.

European Journal of Endocrinology 163 45–54

Introduction

There is widespread anecdotal evidence that GH is misused by athletes for its anabolic and lipolytic properties, despite its use being prohibited by the World Anti-Doping Agency (WADA) (1, 2). The detection of GH abuse in athletes is challenging for a number of reasons: unlike many substances of abuse, it is a naturally occurring substance, and recombinant human GH (rhGH) is identical to the 22 kDa isoform of endogenously produced pituitary GH. Furthermore, serum GH concentrations are highly variable as GH is secreted in a pulsatile fashion, with both exercise and stress acting as potent stimuli for GH secretion. WADA introduced a test at the Athens Olympic Games based on the measurement of GH isoforms (3). This test was subsequently used at the Turin and Beijing games. No sample tested positive during these games, most likely because the test has a short window of opportunity of <24–36 h. The first abnormal analytical finding was detected in February 2010 following a UK Anti-Doping out-of-competition sample taken from Terry Newton, an English Rugby League player, who subsequently admitted to having taken rhGH.

The GH-2000 project proposed a test based on the measurement of the GH-sensitive markers, insulin-like growth factor 1 (IGF1) and type III pro-collagen (P-III-P) (4). Both of these markers rise in response to GH administration in a dose-dependent manner (5, 6). As these markers are largely unaffected by other regulators of GH secretion, such as exercise or injury, and remain elevated beyond the discontinuation of GH,
the test provides a means of detecting exogenous GH administration for a longer period after the last dose of GH (7–9).

Although a single marker could be used, by combining markers in conjunction with gender-specific equations, ‘discriminant functions’, the sensitivity and specificity of the ability to detect GH abuse can be improved compared with single-marker analysis (4).

The results of the GH-2000 project were presented at an IOC workshop in Rome in March 1999 to review critically and assure the quality of the results (10). The conclusion of the workshop was strong support for the methodology but it was felt that several issues needed to be addressed before the test could be fully implemented at an Olympic games. The biggest issue related to potential ethnic variation in the effects of GH as the vast majority of volunteers in the GH-2000 study were white Europeans.

The GH-2004 project was established to address the issues raised by the IOC workshop. The aim of the current study was to determine whether the response to GH in other ethnic groups is similar to the white European amateur athletes in the GH-2000 double-blind placebo-controlled GH administration study (5, 6).

Subjects and methods

Subjects

The study was conducted in the Wellcome Trust Clinical Research Facility (WTCRF) at Southampton General Hospital. The subjects were recruited through a poster campaign at the University of Southampton and local sports facilities. Subjects were eligible to participate if they were aged 18–35 years, and had trained at least twice a week for more than 1 year. Subjects were recruited from non-white European ethnic backgrounds and were asked to self-assign themselves to one of three ethnic groups: i) Indo-Asian (people whose ancestry was either the Indian subcontinent or the Middle East), ii) Afro-Caribbean or iii) Chinese.

Subjects with neoplastic disease, diabetes mellitus, pregnancy and lactation were excluded. Sexually active women were instructed to use a reliable form of contraception during the study. Subjects were also excluded if they had taken part in any clinical study during the previous 12 weeks. All subjects were required to provide written assurance that they had never used performance-enhancing drugs.

The study was approved by the Southampton and South West Hampshire Local Research Ethics committee. All subjects gave written informed consent to their participation, and the study was conducted in accordance with the ethical principles of the Declaration of Helsinki and the guidelines of good clinical practice. The study was regulated by the Research and Development department of Southampton University Hospitals NHS Trust.

A total of 47 healthy volunteers entered the study. Two dropped out for personal reasons. Data were therefore available for 31 men and 14 women (Table 1).

Study design

The subjects were randomly assigned to treatment with placebo or 0.1 IU/kg per day (low dose) or 0.2 IU/kg per day (high dose) rhGH (Saizen, Serono) in a double-blind fashion. Subjects were taught to self-administer GH by daily s.c. injections at bedtime. Placebo preparations were reconstituted in a volume of solvent identical to that of GH preparations and were administered in a similar fashion. Compliance was assessed by direct questioning of the subjects by the GH-2004 medical research fellow (C M).

Treatment protocol and follow-up studies

Before entering the study, a medical history and physical examination were performed for all subjects. Female subjects underwent a pregnancy test.

The study protocol consisted of 28 days of treatment followed by 56 days of washout period. Study visits were scheduled every week during treatment and on days 35, 42 and 84 during the washout period. Resting venous samples were obtained from the antecubital fossa at baseline and at each visit. The samples were allowed to clot, and the serum was separated by centrifugation at 1200 g, at 4 °C. The serum was stored in aliquots at −80 °C until analysis.

Analytical procedures

All samples were analysed in duplicate by the WADA accredited laboratory in London. The laboratory was blinded to the treatment code, which was broken after all the assay results were entered into the database. The blinding procedure meant that samples of any individual were not necessarily assayed within the same assay run.

Serum IGF1 was measured by the DSL-5600 ACTIVE IGF1 IRMA (Diagnostic Systems Laboratories, Inc., Webster, TX, USA). The intra-assay precision was 3.4, 3.0 and 1.5% at 9.4, 55.4 and 263.6 µg/l respectively. The inter-assay precision was 8.2, 1.5 and 3.7% at 10.4, 53.8 and 255.9 µg/l respectively.

Serum P-III-P was measured by a two-stage sandwich RIA (CIS Biointernational; Oris Industries, Gif-Sur-Yvette Cedex, France). The lower and upper limits of detection are 0.1 units and 14 units of P-III-P/ml respectively. The reported intra-assay variability at 0.8, 1.5 and 4.0 U/ml is 2.9, 2.9 and 4.0% respectively. The inter-assay variability at 0.25, 1.5 and 5.6 U/ml is 11.3, 7.8 and 9.3% respectively.
Data analysis

Statistical analysis was performed using the SAS software (SAS Institute, Inc., Cary, NC, USA). All analysis was performed on the log-transformed values of IGF1 and P-III-P. The concentrations of both biomarkers and the GH-2000 scores at each visit day were assessed against the clean observation values (baseline and placebo-treated samples).

The analysis was carried out using a mixed effect model with subject defined as a random effect. This model specification allows the analysis to take into account that individual responses might vary because of inherent individual differences. The mixed model was defined with each time point as a within-subject factor, and treatment, ethnicity and age as between-subject factors. The GH-2000 study showed marked sexual dimorphism in response to GH, and therefore, the genders were analysed and presented separately.

The analysis was performed separately for each treatment arm following an adjustment for the reciprocal of age of each subject. Results are expressed as the mean ± S.E.M.

Adjusting for IGF1 assay differences

In the GH-2000 studies, serum IGF1 was determined by a hydrochloric acid–ethanol extraction RIA (Nichols Institute Diagnostics, San Juan Capistrano, CA, USA) (5). Initially, it was planned that identical assays would be used for the GH-2004 studies but in 2005, Nichols Institute Diagnostics ceased trading and an alternative IGF1 assay was therefore needed. In order to compare the results of the current study with the previous GH-2000 studies, the values of serum IGF1 measured in this study were converted to the scales used by the GH-2000 group.

To adjust for assay differences between the GH-2000 and GH-2004 studies, 73 GH-2000 samples were analysed by the Nichols RIA and the DSL-5600 IRMA as previously described (9). A conversion factor was estimated to convert measured concentrations from the DSL-5600 scales to the measurement scales of the Nichols RIA assay.

\[ \text{GH-2000 RIA} \]

\[ Z_0 = 660 - \frac{C}{255} \]

\[ \text{DSL-5600 IRMA} \]

No adjustment for P-III-P was needed because no significant adjustments had been made to the assay structure and methodology.

The GH-2000 detection method

The previously published GH-2000 discriminant function formulae are as follows (4):

Male score = \(-6.586 + 2.905 \times \log(\text{P-III-P}) + 2.100 \times \log(\text{IGF1}) - 101.737/\text{age} \)

Female score = \(-8.459 + 2.454 \times \log(\text{P-III-P}) + 2.195 \times \log(\text{IGF1}) - 73.666/\text{age} \)

These discriminant formulae were derived from amateur athletes participating in the GH-2000 double-blind placebo-controlled GH administration study (5, 6) and are calibrated against the GH-2000 elite athlete population (11). They are defined such that the mean GH-2000 score is 0 and S.D. is 1 in elite

Table 1 Characteristics of subjects by GH allocation at baseline. Data are mean ± S.E.M.

<table>
<thead>
<tr>
<th></th>
<th>High dose</th>
<th>Low dose</th>
<th>Placebo</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n)</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>0.42</td>
</tr>
<tr>
<td>Age</td>
<td>23.3 ± 0.52</td>
<td>22.6 ± 0.52</td>
<td>24.6 ± 0.52</td>
<td>0.34</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Indo-Asian</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Afro-Caribbean</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Chinese</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>24.2 ± 0.58</td>
<td>23.4 ± 0.58</td>
<td>25.9 ± 0.58</td>
<td></td>
</tr>
<tr>
<td>IGF1 (µg/l)</td>
<td>545 ± 25.0</td>
<td>578 ± 25.0</td>
<td>554 ± 25.0</td>
<td></td>
</tr>
<tr>
<td>P-III-P (U/ml)</td>
<td>0.36 ± 0.02</td>
<td>0.56 ± 0.02</td>
<td>0.40 ± 0.02</td>
<td></td>
</tr>
<tr>
<td>GH-2000 score</td>
<td>-1.7 ± 0.17</td>
<td>-0.9 ± 0.17</td>
<td>-1.2 ± 0.17</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n)</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>0.43</td>
</tr>
<tr>
<td>Age</td>
<td>21.9 ± 0.28</td>
<td>22.1 ± 0.28</td>
<td>23.3 ± 0.28</td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>22.4 ± 0.34</td>
<td>26.7 ± 0.34</td>
<td>23.1 ± 0.34</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Indo-Asian</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Afro-Caribbean</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Chinese</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IGF1 (µg/l)</td>
<td>457 ± 19</td>
<td>470 ± 19.4</td>
<td>555 ± 19.4</td>
<td></td>
</tr>
<tr>
<td>P-III-P (U/ml)</td>
<td>0.34 ± 0.01</td>
<td>0.42 ± 0.01</td>
<td>0.32 ± 0.01</td>
<td></td>
</tr>
<tr>
<td>GH-2000 score</td>
<td>-1.5 ± 0.10</td>
<td>-1.2 ± 0.10</td>
<td>-1.6 ± 0.10</td>
<td></td>
</tr>
</tbody>
</table>

*Asian subjects: 14 of the men were from the Indian subcontinent, and three were Arabic. All of the female Asian subjects were from the Indian subcontinent.
Although the numbers of subjects within any one ethnic group were small, particularly for the women and Chinese men, there was no difference in the mean baseline IGF1, P-III-P or GH-2000 score between the three ethnic groups.

Effect of GH on IGF1

IGF1 concentrations rose in response to both low- and high-dose GH in both men and women (Fig. 1). In men, mean IGF1 concentrations were significantly higher in both GH-treated groups throughout the treatment period (day 28 high-dose GH \( P < 0.0001 \), low-dose GH \( P = 0.041 \) versus placebo). Mean IGF1 remained higher in the high-dose GH group than in placebo for up to 7 days after the discontinuation of GH (\( P = 0.003 \)).

In women, mean IGF1 concentrations were significantly higher in the high-GH group throughout the treatment period compared with placebo (day 28

![Image of IGF1 concentration changes](https://www.eje-online.org)

### Results

#### Baseline characteristics

Fourteen subjects (10 males and 4 females) received the low-dose GH, 16 subjects (11 males and 5 females) received the high-dose GH and 15 subjects (10 males and 5 females) received placebo (Table 1).

The participants in the three study arms were well matched for age and body composition. Baseline IGF1 and P-III-P concentrations were similar in each group. There was no difference in mean baseline GH-2000 score between the groups for either gender. There was no significant correlation between IGF1 and P-III-P levels at baseline (\( P = 0.65 \)). The GH-2000 scores are below zero in all groups reflecting the lower average IGF1 and P-III-P in amateur athletes compared with the elite athletes against whom the GH-2000 score was calibrated (4, 11). This was also found in the original GH-2000 study (5, 6).

![Figure 1](https://www.eje-online.org)
high-dose GH \( P = 0.0019 \). In the low-dose group, mean IGF1 levels were significantly higher than placebo at the visit on day 21 \( (P = 0.03) \) but not on day 28 \( (P = 0.09) \). Mean IGF1 concentrations were not different from placebo in the washout period in the high-dose group.

There was no change in IGF1 concentrations in the placebo group throughout the study period \( (P = 0.48 \text{ for men and } P = 0.23 \text{ for women}) \).

**Effect of GH on P-III-P**

P-III-P concentrations rose in response to both low- and high-dose GH in both men and women (Fig. 2). In men, mean P-III-P concentrations were significantly higher in both GH-treated groups throughout the treatment period (day 28 high-dose and low-dose GH \( P < 0.0001 \) versus placebo). In the high-dose GH group, mean P-III-P remained higher than placebo for up to 56 days after the end of the discontinuation of GH \( (P = 0.01 \text{ on day 84}) \). In the low-dose GH group, mean P-III-P remained higher than placebo for up to 14 days after the end of the discontinuation of GH \( (P = 0.006 \text{ on day 42}) \).

In women, mean P-III-P concentrations were significantly higher in the high GH group from day 14 of the treatment period compared with placebo (day 28 high-dose GH \( P = 0.0001 \)). There was no difference in P-III-P between the high-dose GH group and placebo during the washout period. There was no difference in P-III-P between the low-dose GH group and placebo.

There was no change in P-III-P concentrations in the placebo group throughout the study period.

**Effect of GH on GH-2000 score**

The GH-2000 score increased in response to low- and high-dose GH in both men and women (Fig. 3).

In men, the mean GH-2000 score was significantly higher in GH-treated subjects throughout the treatment period (low- and high-dose GH versus placebo...
Women study in both men (\(P<0.0001\) on visit day 28). Compared with the placebo group, the mean score remained significantly higher in the high-dose GH group throughout the washout period (\(P=0.01\) on day 84) and for 14 days in the low-dose GH group (\(P=0.002\) on day 42).

In women treated with high-dose GH, the mean GH-2000 score was significantly higher than placebo throughout the treatment period but not in the washout phase (\(P<0.0005\) versus placebo on visit day 28). There was no significant difference in GH-2000 score between the low-dose GH group and placebo group throughout the study.

There was no change in GH-2000 score in the placebo-treated group throughout the treatment period (men \(P=0.62\); women \(P=0.49\)).

**Effect of ethnicity within the GH-2004 study**

While the numbers were limited, particularly for women and for the Chinese men, there were no trends suggesting any differences between the ethnic groups in peak IGF1 and P-III-P concentrations, or the GH-2000 score after GH administration.

**Comparison with the GH-2000 study**

The subjects in the GH-2000 study were slightly older than the GH-2004 subjects (men 25.7±0.6 vs 23.5±0.6 years, \(P=0.018\); women 25.6±0.6 vs 22.5±0.5 years, \(P=0.006\)).

In keeping with the younger age of the subjects in the GH-2004 study, the mean assay adjusted baseline and placebo-treated IGF1 concentrations were higher in the GH-2004 study in the men (\(P=0.001\); Table 2). There was no difference in the assay adjusted baseline and placebo-treated IGF1 concentrations in the women (\(P=0.57\)). By contrast, the baseline and placebo-treated P-III-P concentrations were lower in the GH-2004 study in both men (\(P=0.01\)) and women (\(P=0.001\)).

The baseline and placebo-treated GH-2000 score were lower in the GH-2004 study (men \(P=0.005\), women \(P=0.004\)).

Following GH administration, there was no difference in the peak IGF1 or P-III-P between the studies (Figs 4 and 5, Table 2). The peak post-GH GH-2000 score was lower in the men of the GH-2004 study but there was no difference in women (Fig. 6). There was no difference between studies in the maximal change in IGF1, P-III-P and GH-2000 score in response (calculated as difference between peak and baseline) to GH in either men or women. There was no significant effect of ethnicity in response to GH even when both treatment groups were combined.

**Discussion**

The study has confirmed that serum IGF1 and P-III-P concentrations, and GH-2000 score increase in response to GH administration in a dose-dependent manner in healthy male and female amateur athletes of non-white European ethnic origin. The gender differences observed in the GH-2000 study were also seen in the current study (4–6). Although there were baseline differences in IGF1 and P-III-P between the GH-2000 and GH-2004 studies, there was no measurable difference in the peak IGF1 and P-III-P or maximal change in these markers following GH administration. This indicates that ethnicity should not have a major impact on the performance of the GH-2000 detection method. This is critical to the implementation of the GH-2000 test for the detection of GH abuse in professional sportsmen and women, and addresses the concerns raised by the IOC expert panel in Rome.

Previous studies of non-athletic populations have shown that IGF1 concentrations tend to be slightly higher (~10%) in Caucasians of Northern European ancestry (12, 13), although this may reflect differences in body composition (14). Two studies have examined

---

**Table 2** Baseline, delta and peak insulin-like growth factor 1 (IGF1), type III pro-collagen (P-III-P) and GH-2000 scores in the men and women who participated in the GH-2000 and GH-2004 double-blind placebo-controlled GH administration study.

<table>
<thead>
<tr>
<th></th>
<th>GH-2000 study</th>
<th>GH-2004 study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IGF1 (µg/l)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>297±10</td>
<td>361±18</td>
</tr>
<tr>
<td>Baseline</td>
<td>871±40</td>
<td>888±88</td>
</tr>
<tr>
<td>Post-GH</td>
<td>548±191</td>
<td>518±382</td>
</tr>
<tr>
<td>Delta maximal</td>
<td>0.53±0.02</td>
<td>0.43±0.04</td>
</tr>
<tr>
<td>change in IGF1</td>
<td>1.71±0.11</td>
<td>1.49±0.17</td>
</tr>
<tr>
<td>Peak IGF1</td>
<td>1.15±0.52</td>
<td>1.00±0.73</td>
</tr>
<tr>
<td>(µg/l)</td>
<td>−0.7±0.1</td>
<td>−1.3±0.2</td>
</tr>
<tr>
<td>Placebo</td>
<td>0.7±3</td>
<td>3.4±0.4</td>
</tr>
<tr>
<td>Baseline</td>
<td>5.1±1.5</td>
<td>4.7±2.2</td>
</tr>
<tr>
<td>Post-GH</td>
<td>1.4</td>
<td>2.6</td>
</tr>
<tr>
<td>Delta maximal</td>
<td>0.52</td>
<td>0.41</td>
</tr>
<tr>
<td>change in IGF1</td>
<td>−0.6±0.1</td>
<td>−1.4±0.2</td>
</tr>
<tr>
<td>Peak IGF1</td>
<td>4.9±0.2</td>
<td>3.4±1.6</td>
</tr>
<tr>
<td>(µg/l)</td>
<td>5.1±1.5</td>
<td>3.4±1.6</td>
</tr>
<tr>
<td>Placebo</td>
<td>0.57</td>
<td>0.57</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.64</td>
<td>0.74</td>
</tr>
<tr>
<td>Post-GH</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Delta maximal</td>
<td>0.92</td>
<td>0.40</td>
</tr>
<tr>
<td>change in IGF1</td>
<td>0.92</td>
<td>0.40</td>
</tr>
<tr>
<td>Peak IGF1</td>
<td>0.004</td>
<td>0.29</td>
</tr>
<tr>
<td>(µg/l)</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>Placebo</td>
<td>0.01</td>
<td>0.005</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Post-GH</td>
<td>0.41</td>
<td>0.41</td>
</tr>
<tr>
<td>Delta maximal</td>
<td>0.41</td>
<td>0.41</td>
</tr>
<tr>
<td>change in IGF1</td>
<td>0.41</td>
<td>0.41</td>
</tr>
<tr>
<td>Peak IGF1</td>
<td>4.7±2.2</td>
<td>3.2±2.5</td>
</tr>
</tbody>
</table>

---

www.eje-online.org
IGF1 concentrations in elite athletes (15, 16). An Australian study of 699 males and 404 female elite athletes found no significant difference in IGF1 concentrations between African, Asian, Oceanian and Caucasian athletes. The same study also reported that P-III-P was 8.5% higher in Asians compared with Caucasians but overall the contribution of ethnicity to the variation of IGF1 and P-III-P was <2% (15). The second study, which was undertaken by the GH-2004 project and included 242 male and 62 female elite athletes from 50 different nationalities, found small differences in IGF1 and P-III-P concentrations between athletes of different ethnicities but almost all the observations were below the upper 99% prediction limits derived from white European athletes (16).

In keeping with these findings, the baseline and placebo-treated P-III-P values were lower in the GH-2004 study compared with the white European subjects in the GH-2000 study. By contrast, the baseline and placebo-treated IGF1 concentrations were slightly higher in the male GH-2004 subjects, which may reflect their younger age (17). Consequently, the baseline and placebo-treated GH-2000 score was lower in the GH-2004 study. Although this may reflect the ethnic difference between studies, it may also reflect differences in the levels of training between the studies. Although the entry criteria were similar for both studies, recruitment was more challenging for the GH-2004 because this was restricted to non-white European subjects. Consequently, the subjects in the GH-2004 were more difficult to recruit, and may have been less well trained. This same difficulty meant that fewer

![Figure 4](image1.png)

Figure 4: Change in serum IGF1 from baseline to peak IGF1 in the individual subjects who had received GH 0.2 Units/kg per day (high dose), 0.1 Units/kg per day (low dose) or placebo. These data are compared with the previous GH-2000 project which utilised an identical protocol.

![Figure 5](image2.png)

Figure 5: Change in P-III-P from baseline to peak P-III-P in the individual subjects who had received GH 0.2 Units/kg per day (high dose), 0.1 Units/kg per day (low dose) or placebo. These data are compared with the previous GH-2000 project which utilised an identical protocol.
women and Chinese men were recruited than intended, which reduced the power to detect changes between women of different ethnicities and in Chinese men.

Despite these small baseline differences in marker concentrations, there was no measurable difference in the peak IGF1 and P-III-P or maximal change in these markers following GH administration, indicating that people of different ethnicities respond to GH in a similar manner. This is important for the implementation of the GH-2000 methodology as an anti-doping test. For practical reasons, it is not easy to assess accurately ‘ethnicity’ in an anti-doping setting as many athletes have parents or grandparents from differing ethnic backgrounds. Furthermore, if a test was dependent on ethnicity, it may be possible for a defence lawyer to argue that there had been insufficient validation of the test if their client had a rare ethnic background.

In both the GH-2000 and GH-2004 studies, the double-blind rhGH administration studies were undertaken in amateur athletes for ethical reasons. During the statistical analysis of the GH-2000 study, it was found that IGF1, P-III-P and GH-2000 score were significantly higher in professional athletes than in amateur athletes (11). Since the methodology will only be used in elite athletes, this required an adjustment of the cut-off point in the GH-2000 score to accommodate the higher baseline values seen in elite athletes (4). The GH-2000 score was defined such that the mean of the elite athlete population is 0 and each s.d. is 1. In keeping with this difference between amateur and professional athletes, the baseline GH-2000 scores were again lower in both men and women in the current GH-2004 study.

Following GH administration, the GH-2000 score rose significantly in both men and women, to a similar extent to that observed in the original GH-2000 study that included a large majority of white European people. The post-GH GH-2000 score was lower in the men in the current study, possibly reflecting the lower baseline score. It is conceivable that this may affect the performance of the test, but in reality this study is unsuitable for any formal estimation of ‘sensitivity’. We cannot estimate the sensitivity of the test in the anti-doping arena where the doses are thought to be five- to tenfold higher than those used in this study and regimens of GH administration are unknown. Higher doses and longer periods of administration are likely to result in further increases in IGF1 and P-III-P, particularly in women who are less sensitive to the actions of GH. Nevertheless, a high proportion of men taking the higher dose of GH were detected, in whom compliance was thought to be relatively good. For the same reasons, it is not possible to measure ‘window of opportunity’ accurately, but it is interesting to note that one of the men who had received high-dose GH had a score above 3.7 56 days after discontinuing GH.

The GH-2000 test relies on immunoassays that measure IGF1 and P-III-P and, according to the WADA rules, two assays measuring each analyte and recognising different epitopes are needed. Ideally, as recommended by GH-2000 in 1999, the IOC and WADA should own the assays to retain the control of the assay performance and to prevent problems arising when manufacturers of the commercial assays make changes that affect the cut-off level for the GH-2000 score (18). Following the Rome meeting, the Institute for Bioanalytics (IBA, Branford, CT, USA) was successful in obtaining funding from the US anti-doping agency (USADA) to develop two ‘in-house’ IGF1 and P-III-P immunoassays but this work was never completed.

As in-house assays were not available, the GH-2004 project initially planned to use the same commercial assays as the GH-2000 project. In the interim, however, and in keeping with the concerns raised to the IOC in 1999, the Nichols IGF1 RIA was withdrawn from the market (5). An alternative commercial assay was
chosen but in order to interpret the results from the DSL IRMA used in this study, it was necessary to convert the DSL IGF1 values to the scales used in the GH-2000 study. The decision to use the GH-2000 measurement scale as the ‘gold standard’ was made to maintain consistency across the substantial information previously published by the GH-2000 group.

The relationship between the DSL and Nichols assay has been studied by Krebs et al. (19), who found that the DSL assay had a higher concordance with the Nichols assay than four other commercially available IGF1 immunoassays (IDS OCTEIA IGF1, DRG IGF1 600 ELISA, Medignost IGF1 ELISA and Siemens Immulite 2500 IGF1). There was the anticipated spread of data around the fitted line, which adds a degree of measurement uncertainty that must be taken into account when calculating the appropriate cut-offs for the test. The 95th percentile of the dispersion of residuals was 56.9 mg/l, and the correlation coefficient was 0.864. Consistent with our analysis, the median of the DSL IGF1 results was 28.6% higher than Nichols values. In the Bland–Altman analysis, the mean difference was 30.8 μg/l, and the scattering of results appeared homogeneous across the measurement range.

The need to use different assays will have reduced the precision with which we can estimate differences in mean response to GH between the white European recreational athletes of the GH-2000 study and non-Caucasian athletes of the current study. It is reassuring, however, that the results obtained were comparable to the original GH-2000 study. Nevertheless, before the test can be finally implemented, WADA has asked that further validation of another set of commercial assays in a population of elite athletes be undertaken.

This study has a number of limitations; it did not examine the ethnic differences in responses over the entire treatment and washout periods because the difference in the timing of the samples between the GH-2000 and GH-2004 studies precluded area under the curve analysis; however, this is less relevant to antidoping than peak response. The small numbers in each individual ethnic group meant that the power to detect differences between the three ethnicities studied was limited. Not all the individuals attended for all their scheduled follow-up visits; this was particularly apparent on day 35 which limited our ability to ascertain how quickly the markers and scores fell once rhGH was discontinued. In this study, the samples from one individual were not all assayed in the same assay run in order to maintain blinding in the laboratory as this reflects current anti-doping procedures; however, this approach may have reduced the power to detect differences between individuals because of inter-assay variability. Larger studies involving more amateur athletes would be needed to prove conclusively that there are no differences in the response to GH between different ethnic groups.

In conclusion, this study has shown that there is no evidence that non-Caucasian athletes respond differently to GH than white European athletes, at least in terms of the peak and maximal change in IGF1 and P-III-P. While the study cannot exclude small differences between ethnic groups, these data indicate that ethnicity should not have a major impact on the performance of the GH-2000 detection method.

Declarations of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

The GH-2004 study was funded by the World Anti-Doping Agency and US Anti-Doping Agency without which this study would not have been possible.

Acknowledgements

We would like to thank Clare Hartley and Rod Park, the GH-2004 project managers for their administrative support for the study and all the subjects who volunteered for the study. We would like to thank the nurses of the Wellcome Trust Clinical Research Facility at Southampton University Hospitals Trust for their assistance in recruitment and follow-up of the GH-2004 subjects. We thank the University of Southampton sports clubs for allowing us to approach potential participants on campus. We also thank Serono for providing the GH at a reduced cost and for organizing the randomisation for the study. We are indebted to the GH-2000 Team who left us a legacy of invaluable samples that have enabled us to align currently available assays with those in use 10 years ago and to develop a methodology for aligning new assays in the future.

References


www.eje-online.org


17 Toogood AA. Growth hormone (GH) status and body composition in normal ageing and in elderly adults with GH deficiency. *Hormone Research* 2003 60 (Supplement 1) 105–111.
