Gender has to be taken into account in diagnosing adult growth hormone deficiency by the GHRH plus arginine test

Helene M. Markkanen\textsuperscript{a,}*, Tuula Pekkarinen\textsuperscript{b}, Esa Hämäläinen\textsuperscript{a,c}, Matti J. Välimäki\textsuperscript{b}, Henrik Alftahan\textsuperscript{c}, Ulf-Håkan Stenman\textsuperscript{c}

\textsuperscript{a} Helsinki University Central Hospital, HUSLAB, Helsinki, Finland
\textsuperscript{b} HUS Abdominal Centre, Endocrinology, Helsinki University Hospital, Helsinki, Finland
\textsuperscript{c} Department of Clinical Chemistry, Helsinki University Central Hospital, Helsinki, Finland

\textbf{A B S T R A C T}

\textbf{Objective:} Data on the effect of gender on the interpretation of the GHRH plus arginine stimulation test (GHRH + ARG test) is controversial. We validated the GHRH + ARG stimulation test in control subjects and patients with organic or idiopathic pituitary disease and a suspicion of adult growth hormone deficiency (AGHD) using the Immulite 2000 XPI GH assay.

\textbf{Design:} We studied 126 apparently healthy adults (median age 38.8 years) and 34 patients with a suspicion of AGHD (median age 42.2 years). Identification of AGHD with the GHRH + ARG test was investigated with commonly accepted BMI-related consensus cut-off limits for peak GH concentrations. Serum samples collected during the GHRH + ARG test were analysed for GH in 2014–2015. Serum IGF-1 concentrations were studied as a reference.

\textbf{Results:} In 14 of 65 (22\%) control males the GH peak value was below the BMI-related cut-off limits for GH deficiency indicating a false diagnosis of AGHD. All control females had a normal GHRH + ARG response. Median peak GH response was significantly (\(p < 0.001\)) higher in female (39.3 \(\mu\)g/L) than in male controls (21 \(\mu\)g/L). According to consensus cut-offs all but one young female patient had a deficient response compatible with a diagnosis of AGHD.

\textbf{Conclusions:} The GH response to stimulation by GHRH + ARG is gender-dependent, being lower in healthy males than in females. Gender should be considered when defining cut-off limits for peak GH concentrations in the GHRH + ARG test. The presently used BMI-related cut-off levels will lead to a significant misclassification of males as GH deficient.

\textbf{1. Introduction}

Adult growth hormone deficiency (AGHD) with evidence of hypothalamic-pituitary disease is recognized as a clinical entity characterized by increased abdominal fat mass, decreased muscle mass, lowered bone density and adverse effects on quality of life and cardiovascular morbidity [1–3]. However, these signs and symptoms are nonspecific, and accurate diagnosis based on laboratory tests is needed for successful AGHD therapy.

The diagnosis of AGHD is based on measurement of serum GH in response to pharmacological stimulation in patients with symptoms suggestive of AGHD [4,5]. In Europe one of the preferred tests is the GHRH + ARG test. It is well tolerated and reproducible [6–8]. Cut-off values for the GHRH + ARG test may vary based on the controls used [8,9]. The presently used consensus cut-off criteria for peak GH concentrations in the GHRH + ARG test are mainly based on the study of Corneli et al., in which the effect of gender was not addressed, and thus the criteria are only BMI-specific [4,5]. The reliability and reproducibility of pharmacological stimulation tests in the diagnosis of GHD are still under discussion. Albeit findings have been somewhat controversial, age and gender obviously also affect the GH response in the GHRH + ARG test [6,8,11–14]. Furthermore, there is a significant variation between GH results obtained by assays from different manufacturers, which invalidates the use of common cut-off limits [15]. There is the high risk of misclassification when using generally accepted consensus cut-offs [16].

The purpose of this study was to validate the GHRH + ARG test and it's cut-off limits for diagnosis of AGHD using the GH Immulite 2000 XPI assay calibrated against the WHO standard IS 98/574. A specific aim was to study BMI-related consensus cut-off values in the GHRH + ARG test.

\textsuperscript{*} Corresponding author at: HUSLAB, P.O. Box 720, 00029 HUS Helsinki, Finland.
E-mail address: helene.markkanen@hus.fi (H.M. Markkanen).
test.

2. Material

2.1. Control subjects

We recruited 119 apparently healthy controls (59 males and 60 females) through advertisement in our hospital newspaper. The controls consisted of medical students, hospital personnel and their relatives. To be eligible, participants were required to be healthy and without any symptoms indicative of disease. Exclusion criteria were pregnancy and a known or suspected disease or any symptoms or signs indicating disease or medication.

Our aim was to recruit ten females and ten males aged 20–30 years, 30–40 years, 40–50 years and 50–60 years. Enough healthy controls were not found in the oldest group. Among the controls, five used statins, one beta blocker, one tamsulosin and one thyroxine for hypothyroidism because of thyroidectomy and vitamin B12 injections for pernicious anaemia. No one used estrogens. We also included seven subjects who had been referred due to fatigue but without history of pituitary or hypothalamic disease, as controls. They had been studied by arginine (L-arginine monohydrochloride, Braun, Melsungen, Germany) 0.5 g/kg (maximally 30 g) over 30 min. Blood samples were carefully collected without abnormal results. They had been studied in subjects who had been referred due to fatigue but without history of pituitary or hypothalamic disease. As controls, they had been studied with no abnormal findings in hormone concentrations and other laboratory tests; two of them had had pituitary imaging with normal findings. One male had type 1 diabetes and used insulin, and another used testosterone for primary hypogonadism.

2.2. Patients with suspected GH deficiency

We studied 34 patients (20 males and 14 females) with pituitary disease and suspicion of GH deficiency. Of these, 24 had undergone pituitary surgery and nine pituitary irradiation. Of our patients six (five males, one female) had received radiation over ten years before testing, and three patients (two male, one female) within ten years. The primary diseases were: eight non-functioning pituitary adenomas, five secreting adenomas, three craniopharyngiomas, three cysts of the Rathke pouch, three hypophysitis, two meningomas, one astrocytoma, one glioma, one ependymoma, one myxomatous pseudotumor, one histiocytosis, one Sheehan syndrome, one congenital panhypopituitarism, two idiopathic GH deficiency diagnosed in childhood (one with continuing secondary hypogonadism) and one secondary hypogonadism. Of these patients 20 had three, 13 had two, one had one anterior pituitary deficiency other than GH, and 11 patients had hypothalamic diabetes insipidus. The patients were on adequate replacement therapy of other deficiencies during the GHRH + ARG test. Of the patients, 12 used oral estrogen and one used transdermal estrogen.

The study was approved by the Ethical Committee of Helsinki University Hospital, and a written informed consent was obtained from all subjects.

3. Methods

The subjects arrived at the hospital after an overnight fast and refrained from strenuous exercise in the morning of the test. Women were studied during the follicular phase of the menstrual cycle (days 7–11). The GHRH + ARG stimulation test was started at 7.00–10.30 (median 7.30) a.m. Intravenous cannulas (i.v.) were inserted in each arm for blood sampling and infusions. One μg/kg GHRH (GHRH(1–29), GEREF Serono, Italy) was administered as an i.v. bolus at time 0 min, followed by arginine (L-arginine monohydrochloride, Braun, Melsungen, Germany) 0.5 g/kg (maximally 30 g) over 30 min. Blood samples were drawn at ~15, 0, 15, 30, 45, 60, 75 and 90 min.

Serum was separated by centrifugation and duplicate tubes were stored frozen for 4–6 weeks at 20 °C and then at ~80 °C until analysed.

During the initial phase of the study, during years 2001–2008, serum samples were analysed by a time-resolved immunofluorometric GH assay (AutoDELFIA, PerkinElmer, Wallac, Turku, Finland), which was our routine method at that time. A duplicate sample from the GHRH + ARG test was saved at ~80 °C. The basal and peak samples were reanalysed with the Immulite 2000 XPi in 2014–2015. Thus these samples had been frozen and thawed once before analysis. In order to evaluate the possible effect of storage on GH levels, we reanalysed 67 samples from 2001 to 2003 in 2013 by the AutoDELFIA assay. This confirmed that there was no loss of GH during storage.

The Immulite 2000 XPi GH assay (Siemens, Healthcare Diagnostics, Los Angeles, CA, USA) is an immunochemiluminometric assay calibrated against the WHO International Standard (IS) 98/574. It recognizes both 22-kDa and 20-kDa hGH. The lowest reportable concentration was 0.05 μg/L. For statistical analyses samples with lower concentrations were assigned a value of 0.025 μg/L.

The GH concentrations in basal and peak samples determined with the Immulite 2000 XPi (n = 272) correlated strongly with the AutoDELFIA results (r = 0.997, p < 0.001). The Passing-Bablok regression equation for Immulite (y) vs. AutoDELFIA (x) was y = 1.023 * x + 0.01. For diagnosis of AGHD we used BMI-specific cutoff limits (11.5 μg/L, BMI < 25 kg/m2; 8.0 μg/L, 25–30 kg/m2; 4.2 μg/L, BMI > 30 kg/m2), which have been reported in the consensus statement [4,5].

Serum IGF-1 was measured during the same thawing with the Immulite 2000 XPi. IGF-1 assay that has been calibrated against the WHO International Reference Reagent (IRR) 87/518. IGF-1 concentrations were compared with age-specific reference values provided by the assay manufacturer.

3.1. Statistical methods

The results are expressed as median and range or mean and standard deviation as appropriate. The non-parametric Mann-Whitney U test was used to compare non-normally distributed continuous variables between groups and two-sample t-test to compare normally distributed variables between groups. Spearman correlation was used to test associations between continuous variables. The difference in peak GH, basal GH and IGF-1 values between groups was studied using analysis of covariance after adjustment for potential confounding factors, gender, age and BMI. Age and BMI adjusted gender difference within groups was evaluated using analysis of covariance. Due to the positively skewed distributions, log-transformed peak GH, basal GH and IGF-1 values were used in analysis of covariance. p-Values < 0.05 were considered statistically significant. SPSS for Windows version, 23.0 (IBM Corp. Armonk, NY, USA) was used for all statistical analyses.

4. Results

Table 1 shows characteristics of the study population. Among control subjects there was no significant difference in age between genders, but in males BMI was significantly higher than in females (p < 0.01). Table 2 shows basal and peak serum GH and IGF-1 concentrations. Among the controls, females had both higher basal and peak GH level than males, and this gender difference remained significant after adjustment for age and BMI (p < 0.001). The peak GH value correlated negatively with BMI (r = −0.56; p < 0.001) and age (r = −0.37; p < 0.001).

All 61 female controls but only 51 of 65 males (78.5%) had peak GH values above the BMI-related cut-off limits for GH sufficiency (Fig. 1). Of those 14 males, who were classified as GH insufficient, three had a BMI below 25 kg/m2, ten a BMI of 25–30 kg/m2 and one over 30 kg/m2 (Fig. 1). BMI was significantly higher in these than in the other male controls (mean BMI 26.9 kg/m2 vs. 24.6 kg/m2 respectively, p < 0.05). The median age of these males was 40 year (range 28.1–59.4 year) and there was no difference in age compared with the other male controls. Serum IGF-1 concentrations in these 14 males were within the reference ranges established by the manufacturer of the Immulite assay.
Among patients, females had a higher peak GH than males (median 2.2 μg/L vs. 1.4 μg/L; p < 0.01), but the difference in peak GH between genders did not remain significant after adjustment for age and BMI (p = 0.15) (Table 2). Basal GH was also higher in females than in males (median 0.12 μg/L vs. 0.03 μg/L; p < 0.05) and the difference remained significant after adjustment for age and BMI. Among patients, peak GH correlated negatively with age (r = –0.43; p < 0.05), but not with BMI (r = 0.09; p = 0.63). Using current BMI-related cut-off limits for diagnosis of AGHD, a 19-year-old female patient with a BMI of 33.9 kg/m² had a peak GH response of 16.3 μg/L and was thus classified as “GH sufficient”.

The peak GH response was lower in patients than in controls (median 1.5 μg/L; range 0.03–16.3 μg/L vs. 27.2 μg/L; 2.7–116.0 μg/L, p < 0.001) (Table 2 and Fig. 1). Two male patients (40 year; 2.8 μg/L and 44 year; 2.9 μg/L) and the above-mentioned young female had a peak GH above the lowest value of controls (2.68 μg/L for males, 11.9 μg/L for females) (Fig. 1). All three patients had a BMI above 20 kg/m². Basal GH and IGF-1 levels were also lower in patients than in controls (p < 0.01) (Table 2). The differences between groups remained significant after adjustment for gender, age and BMI for peak (p < 0.001) and basal GH (p < 0.001). The difference in IGF-1 concentrations was also significant (p < 0.001).

There was no gender difference in serum IGF-1 concentrations in controls or patients either before or after adjustment for age and BMI (p = 0.20 vs. p = 0.16, respectively) (Table 2). Serum IGF-1 correlated with peak GH both in controls (r = 0.20; p < 0.05) and patients (r = 0.60; p < 0.001), but not with a basal GH (r = 0.14; p = 0.13 and r = 0.09; p = 0.64, respectively). The IGF-1 concentrations correlated with BMI in patients (r = 0.48; p < 0.01), but not in controls (r = –0.06; p = 0.52). The serum IGF-1 concentrations were below the age-related reference values in 25 of 34 (73.5%) patients and the rest were within the lowest third of the reference range. Among 126 controls, six had a value below and three above the reference range.

### Table 1

<table>
<thead>
<tr>
<th>Healthy controls</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>(n = 126)</td>
</tr>
<tr>
<td>Age (year)</td>
<td>39.1 (11.3)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>73.7 (13.0)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>24.3 (3.1)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>83 (66%)</td>
</tr>
<tr>
<td>Male</td>
<td>37 (29%)</td>
</tr>
</tbody>
</table>

The statistical significance for the difference between healthy controls and patients (p < 0.05, *p < 0.01, **p < 0.001) and for the difference between males and females within control and patient groups (*p < 0.05, **p < 0.01, ***p < 0.001).

### Table 2

<table>
<thead>
<tr>
<th>N</th>
<th>Healthy controls</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All n = 126</td>
<td>Males n = 65</td>
</tr>
<tr>
<td>Peak GH (μg/L)</td>
<td>27.2</td>
<td>21.0 (2.68-71.0)</td>
</tr>
<tr>
<td>Basal GH (μg/L)</td>
<td>0.23</td>
<td>0.09 (0.03-1.35)</td>
</tr>
<tr>
<td>IGF-1 (nmol/L)</td>
<td>21.3</td>
<td>20.9 (9.72-42.3)</td>
</tr>
</tbody>
</table>

*p < 0.05, **p < 0.01; the statistical significance for the difference between control and patient groups using analysis of covariance after adjustment for age and BMI.

### 5. Discussion

In accordance with a recent study by Deutschbein et al. [14] we found that the peak GH values in the GHRH + ARG test are clearly higher in healthy females than in males within the same BMI group. With the current BMI-related cut-off limits of AGHD, all female controls were classified as GH sufficient, whereas 22% (14/65) of the males were misclassified as insufficient. Thus, if this gender difference is not taken into account, which is the case in the present guidelines of AGHD [4,5] men are at risk of being falsely diagnosed as GH deficient.

A gender difference has been demonstrated in at least three previous studies [8,12,17] with a smaller number of healthy subjects than in the present study. In these, the significance of the gender difference for diagnosis of AGHD was not emphasized. So far, gender-related cut-off limits are not in routine use. Estrogens play a role in mediating the gender difference in GH secretion. Pulsatile GH levels have been reported to be higher in the late follicular and mid-luteal phase than in the early follicular phase [18]. A strength of our study was that all female controls were studied during the follicular phase of the menstrual cycle (days 7–11) and no one used estrogen.

All patients, including three patients who had received radiation therapy within 10 years [19], but not a 19.3 years-old female with a peak GH of 16.3 μg/L, had a GH response below the present BMI-specific cut-off limits for adult GHD. Serum IGF-1 of this patient was below the lower reference limit. This patient had been operated for a cranioopharyngioma and received substitution therapy; thyroxine, hydrocortisone, oral estrogen and desmopressin. Oral estrogen treatment is known to reduce serum IGF-1 concentration and increase the GH concentration [20]. Based on findings of high GH concentrations during puberty, it has been suggested, that higher cut-off limits (15.1–20.3 μg/L) should be used for diagnosis of GHD [21-23] during a transition period between 16 and 25 years. However, the results concerning the influence of age on the peak GH in the GHRH + ARG test are diverging.
earlier it was thought that age has no effect [6,24] but Colao et al. proposed that the cut-off values should be based on BMI and age. The authors suggested the following cut-off limits for subjects aged 15–25 years: 15.6 µg/L for a BMI < 25 kg/m², 11.7 µg/L for 25–30 kg/m² and 8.5 µg/L for over 30 kg/m² [13]. In that study, there was no difference in peak GH between males and females (mean 41.2 µg/L vs. 42.6 µg/L, respectively).

Although the correlation between GH assays is generally good, there are significant differences between some assays. Müller et al. [15] evaluated commercially available GH immunoassays, which all except the AutoDELFIa (IS 80/505) and BC-IRMA (IS 88/624), were calibrated against the second International Standard for GH, WHO IS 98/574, which is advocated in the recent consensus statement [25]. Considerable between-method differences were observed in mean concentrations of 312 serum samples: Siemens, Immulite 2000, (5.90 µg/L); PerkinElmer, AutoDELFIa, (5.62 µg/L); IDS, iSYS (5.28 µg/L); DiaSorin, Liaison (7.46 µg/L); Mediagnost, ELISA (3.91 µg/L); Beckman Coulter, UniCel Dxl 800 Access (3.91 µg/L) and Beckman Coulter, BC-IRMA (3.19 µg/L) [15]. Thus, the interpretation of GH stimulation results is dependent on the assay used. Our method, Immulite 2000, gave the highest concentration. Assays recognizing both 22 kDa and 20 kDa hGH forms are expected to give higher concentrations than methods specific for 22 kDa GH [26].

Most patients (74%) had subnormal serum IGF-1 concentrations and the rest of the patients had IGF-1 within the lowest third of the age-related reference values. The biological within-subject variation for IGF-1 is high (CV 20%) and a single serum IGF-1 measurement is not necessarily enough [27]. There is a remarkable overlap between patients with AGHD and healthy subjects, especially in subjects over 40 years of age. Thus, a normal IGF-1 level does not rule out GHD [4,28]. Furthermore, despite overtly suppressed GH secretion, serum IGFI-1 levels are often within the reference range for healthy obese subjects [4]. Interestingly, despite a clear gender difference in the GH response to GHRH + ARG, the IGF-1 levels were similar in healthy females and males in the present study.

A strength of our study is the relatively large number of control subjects and well-characterized male and female patients. A limitation is the rather small number of healthy males and females with a high BMI.

All female controls were classified as GH sufficient using the current commonly accepted BMI-related cut-off limits, but for males those cut-offs with Immulite 2000 XPI were too high. The higher BMI of males is not the reason, because the difference remained after using analysis of covariance after adjustment for age and BMI. Unfortunately the BMI groups of males are too small to establish exact cut-offs. In the group of normal BMI (n = 36), the cut-off is by using 2.5 percentile 9.4 µg/L instead of 11.5 µg/L and in the overweight group (n = 26) it is about 3.5 µg/L not 8 µg/L. The obese group is very small (n = 3), according to those results, it is around 2.7 µg/L.

In conclusion, our study shows that in addition to BMI, gender has to be taken into account when interpreting the GHRH + ARG test and it may be one significant and independent determinant of GH peak response. Lower cut-off limits for peak GH concentrations have to be used for males than for females. This may be the case also for other stimulation tests e.g. the arginine alone test [17]. The presently used BMI-related consensus cut-off levels are suitable for females, but will lead to a significant misclassification of males to GH deficient.

Conflicts of interest
None.

References

Fig. 1. Peak serum GH concentrations in male and female controls and patients according to the BMI-related cut-off limits (dotted lines). Solid lines represent the medians for each group. (a) BMI below 25 kg/m², (b) BMI 25–30 kg/m² and (c) BMI over 30 kg/m².


