Impact of different cut-off limits of peak GH after GHRH-arginine stimulatory test, single IGF1 measurement, or their combination in identifying adult patients with GH deficiency

Fausto Bogazzi, Luca Manetti, Martina Lombardi, Clara Giovannetti, Valentina Raffaelli, Claudio Urbani, Ilaria Scattina, Pasquale Pepe¹, Aldo Iannelli², Enio Martino and Giuseppe Rossi¹

Department of Endocrinology and Metabolism, University of Pisa, Ospedale Cisanello, Via Paradisa 2, 56124 Pisa, Italy, ¹Epidemiology and Biostatistics Unit, Institute of Clinical Physiology, National Research Council, Via Moruzzi, 56124 Pisa, Italy and ²Department of Neuroscience, University of Pisa, 56124 Pisa, Italy

(Correspondence should be addressed to F Bogazzi; Email: fausto.bogazzi@med.unipi.it; G Rossi; Email: giuseppe.rossi@ifc.cnr.it)

Abstract

Objective: To evaluate the impact of different peak GH cut-off limits after GHRH-Arg test, IGF1 measurement, or their combination in identifying patients with GH deficit (GHD).

Design and patients: Totally, 894 normal subjects (used for determining IGF1 normative limits) and 302 patients with suspected GHD were included. Different peak GH cut-off limits (used by European (depending on body mass index (BMI)) or North American (4.1 µg/l) Endocrine Societies, by HypoCCs (2.5 µg/l), or with 95% specificity (based on BMI), Method 1, 2, 3, or 4 respectively) and IGF1 were considered.

Methods: Peak GH after GHRH-Arg and IGF1.

Results: Different peak GH cut-off limits recognized different proportions of GHD (range, 24.8–62.9%). Methods 1 and 2 with high sensitivity recognized a higher proportion (95.5 and 92.5% respectively) of GHD among patients with three (T) pituitary hormone deficits (HD), whereas Method 4 (with high specificity) identified 96.7% normal subjects among those without pituitary HD: on the contrary, Method 4 identified 96.7% normal subjects among those without HD; on the contrary, Method 4 identified 96.7% normal subjects among those without HD. On the other hand, Method 4 identified 75% GHD among subjects without HD. Of the total patients, 82% with THD and 84.5% without HD were recognized as GHD or normal respectively by IGF1. Among the remaining patients with THD and normal IGF1, 75% was recognized as GHD by Method 1; among patients without HD and abnormal IGF1, 87.5% was identified as normal by Method 4. Overall, combination of IGF1 and Method 1 or Method 4 identified 95.5% GHD among patients with THD and 98.1% normal subjects among those without HD.

Conclusions: Single peak GH cut-offs have limits to sharply differentiate GHD from normal subjects; IGF1 may be used for selecting patients to be submitted to the GHRH-Arg test; the peak GH cut-off limits to be used for identifying healthy or diseased patients depend mainly on the clinical context.

European Journal of Endocrinology 164 685–693

Introduction

Diagnosis of GH deficiency (GHD) is currently based on the demonstration of impaired GH secretion under stimulatory tests (1–8). GHRH associated with arginine (GHRH-Arg) is the most used test so far (1, 2, 4–8); it is considered equivalent to the insulin tolerance test but without uncomfortable side effects or harmful complications (5, 6). However, which peak GH cut-off limits after GHRH-Arg test should be used for identifying GHD patients is still a matter of argument: in fact, the Endocrine Society has adopted a 4.1 µg/l peak GH (6), whereas the GH Research Society in association with other Medical Societies, including the European Society of Endocrinology (2, 7), suggests peak GH cut-off limits ranging from 4.2 to 11.5 µg/l based on body mass index (BMI) (9). In addition, either a peak GH < 3 µg/l, or < 2.5 µg/l, has been suggested for diagnosing GHD when different stimulatory tests or different methods for measuring GH were used (10–12).

Although high or very low serum IGF1 concentrations almost invariably identify acromegalic or GHD patients respectively (5, 10, 13, 14), normal values are considered not sufficient to exclude the diagnosis of adult GHD (5); reasons for this discrepancy may be due to age, genetics, nutritional status, diseases, and the lack of accurate normative values for IGF1; owing to those findings, IGF1 was found to have 40% sensitivity in a study (15), being considered not always suitable for an appropriate diagnosis of GHD (5).

The objective of this study was to evaluate the role of different peak GH cut-off limits after GHRH-Arg stimulatory test, a single IGF1 measurement or their combination in identifying adult patients with GHD.
Subjects and methods

Study population

All patients and controls were evaluated at the Department of Endocrinology at the University of Pisa.

Healthy subjects

Totally, 849 consecutive healthy subjects (75% females, mean age 47.5 ± 13.4 years) were included in the study for determining the normal limits of serum IGF1 concentrations; diseases (hypopituitarism, acromegaly, malnutrition, diabetes mellitus, renal or liver diseases, malignancy, or other endocrine diseases) as well as therapies (i.e. estrogens) known to affect serum IGF1 concentrations were carefully excluded by medical history, physical examination, or laboratory tests.

Patients with GH excess

A total of 108 patients with active acromegaly (13, 14) (56% females, mean age 51.5 ± 12.5 years) were used for evaluating the discriminatory power of the upper limits of normal range of IGF1.

Patients with suspected adult GHD

A total of 302 consecutive patients with various pituitary diseases (Table 1) were evaluated for adult GHD (60% females, mean age 47.4 ± 15.0 years) using the GHRH-Arg test (16). Diagnosis of adult GHD was made according to the following methods:

Method 1: for patients with BMI < 25 kg/m², a peak of GH ≤ 11.5 μg/l; for BMI 25–30 kg/m², a peak of GH ≤ 8 μg/l; and for BMI > 30 kg/m², a peak of GH ≤ 4.2 μg/l (7); these limits have been adopted by European, Japanese, and American Endocrine Societies and the European, Australian, and American Associations of Clinical Endocrinologists (2, 7, 17).

Method 2: a 4.1 μg/l peak GH cut-off limit without BMI correction, as adopted by the Endocrine Society (6, 9).

Method 3: a ≤ 2.5 μg/l peak GH as proposed by Hartman et al. (10).

Method 4: for patients with a BMI < 25 kg/m², a peak of GH ≤ 2.3 μg/l; for BMI 25–30 kg/m², a peak of GH ≤ 1.5 μg/l; and for BMI > 30 kg/m², a peak of GH ≤ 1.0 μg/l. These cut-off limits were identified by Corneli et al. (17) to achieve a 95% diagnostic specificity.

Before testing for GHD, all patients were under stable replacement therapy with L-thyroxine (1.0–1.6 μg/kg per day), cortisone acetate (25–37.5 mg/day), intranasal desmopressin (5–20 μg/day), testosterone enanthate (250 mg i.m. every 28 days) in men, and transdermal or oral (n = 7 patients) estrogens associated with progesterone in premenopausal women, depending on hormonal status of each patient. Replacement hormone therapy was evaluated through clinical and laboratory tests every 6 months. No patient was treated with GH therapy before enrollment in the study nor received GH therapy during childhood. Patients with previous acromegaly, active Cushing’s disease, diabetes mellitus, malignancy, and renal or hepatic dysfunction were not included in this study. Serum IGF1 concentrations of each patient under oral estrogen replacement therapy were individually evaluated before inclusion in a global statistical analysis; those with abnormal values (n = 1) were excluded from the analysis.

Number of pituitary hormone deficits

Patients screened for adult GHD with the GHRH-Arg test (n = 302) were also stratified keeping into account the number of pituitary hormone deficits except than for GH as follows: total (three deficits), partial (one to two deficits), or absent (0 deficit) hypopituitarism, in agreement with previous reports and with the GH Research Society consensus (1, 10, 18). Classification of

Table 1

<table>
<thead>
<tr>
<th>Pathology</th>
<th>No. of patients (%)</th>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
<th>Method 4</th>
<th>IGF1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pituitary adenoma</td>
<td>166 (54.9)</td>
<td>114 (68.7)</td>
<td>95 (57.2)</td>
<td>68 (41.0)</td>
<td>50 (30.1)</td>
<td>60 (36.1)</td>
</tr>
<tr>
<td>Other tumors</td>
<td>33 (10.9)</td>
<td>28 (84.8)</td>
<td>25 (75.8)</td>
<td>19 (57.6)</td>
<td>12 (36.4)</td>
<td>23 (69.7)</td>
</tr>
<tr>
<td>Empty sella</td>
<td>41 (13.5)</td>
<td>18 (43.9)</td>
<td>15 (36.6)</td>
<td>10 (24.4)</td>
<td>5 (12.2)</td>
<td>15 (36.6)</td>
</tr>
<tr>
<td>Brain trauma</td>
<td>30 (9.9)</td>
<td>14 (46.7)</td>
<td>12 (40.0)</td>
<td>6 (20.0)</td>
<td>3 (10.0)</td>
<td>3 (10.0)</td>
</tr>
<tr>
<td>Pituitary hypoplasia</td>
<td>5 (1.6)</td>
<td>5 (100)</td>
<td>4 (80.0)</td>
<td>4 (80.0)</td>
<td>2 (40.0)</td>
<td>5 (100.0)</td>
</tr>
<tr>
<td>Idiopathic</td>
<td>8 (2.6)</td>
<td>5 (62.5)</td>
<td>3 (37.5)</td>
<td>3 (37.5)</td>
<td>2 (25.0)</td>
<td>3 (37.5)</td>
</tr>
<tr>
<td>Others</td>
<td>19 (6.2)</td>
<td>6 (31.6)</td>
<td>4 (21.0)</td>
<td>3 (15.8)</td>
<td>1 (5.3)</td>
<td>2 (10.5)</td>
</tr>
<tr>
<td>Total (%)</td>
<td>302</td>
<td>190 (62.9)</td>
<td>156 (51.3)</td>
<td>119 (37.4)</td>
<td>75 (24.8)</td>
<td>111 (36.6)</td>
</tr>
</tbody>
</table>

www.eje-online.org
patients based on the extension of hypopituitarism is currently employed for identifying patients who are considered GHD (1, 10, 18); distinction of GHD patients from non-GHD patients was the basis for the creation of all peak GH cut-off limits, mostly of which derived from the analysis of receiver operating characteristic (ROC) curves (9, 10, 16, 17, 19); usually, for that analysis, two populations are necessary: normal subjects and GHD patients; the latter were invariably chosen among those with three (17) or sometimes with two pituitary hormone deficits (9) except GH. We followed those criteria and considered as GHD patients with total hypopituitarism and as likely normal those without pituitary hormone deficit. However, it cannot be excluded a priori that a proportion of patients with three deficits were not GHD and that, on the contrary, a proportion of patients without pituitary hormone deficit was indeed GHD.

**Assays**

Blood samples were drawn in the morning (starting at 0800 h) on fasting conditions. All samples were collected and analyzed at the Department of Endocrinology at the University of Pisa. GH was measured by automated Advantage chemiluminescent GH assay (DiaSorin SpA, Saluggia, Italy) and IGF1 by an automated RIA (DIAsource ImmunoAssays S.A., Nivelles, Belgium), which included ethanol extraction. The sensitivity of GH assay was 0.05 μg/l; the intra- and inter-assay coefficients of variation were 4.1 and 7.3%, respectively. The sensitivity of IGF1 assay was 7 μg/l; the intra- and inter-assay coefficients of variation were 4.7%, 2.1%, 1.9%, and 5.5, 4.1, and 4.9% for low, medium, and high points of the standard curve respectively. The assay was calibrated against the WHO International Reference Reagent (IRR) 87/518. Normal values for serum IGF1 were determined in this study (Table 2). The GHRH-Arg (GHRH 1–29, (GEREF, Serono), 1 μg/kg i.v. at 0 min, arginine hydrochloride, 0.5 g/kg i.v. over 30 min from 0 to +30 min, up to a maximum of 30 g) was chosen because most peak GH cut-off limits have been obtained using this test (1, 2, 6, 7, 16, 20) and carried out as previously reported (16). The peak GH was the highest value of serum GH over the 90 min period of the GHRH-Arg test.

**Normal range for IGF1**

A total of 849 normal subjects were divided in 5-year periods from 20 to 79 years. Normal limits of serum IGF1 were established using three different statistical methods, which gave similar results (Table 2). The regression method was finally chosen for the absence of inconsistencies among the different age classes both for the lower and the upper limits. As expected, serum IGF1 concentrations progressively decreased with the age, with which were negatively correlated ($r = -0.63$, $P < 0.001$). There was a weak relation between IGF1 and BMI ($r = -0.20$, $P < 0.001$); however, BMI lack any relationship with IGF1 when adjusted for age (partial $r = -0.03$, $P = 0.389$); this effect was due to the relationship between age and BMI ($r = 0.26$, $P < 0.001$). The discriminatory capability of the upper limit of normal IGF1 levels to identify patients with active acromegaly was 100% (95% confidence interval (CI), 97–100).

**Statistical analysis**

Continuous variables were expressed as mean ± s.d. Qualitative variable were expressed as percentage. The normal limits of serum IGF1, based on the healthy subjects, were computed for each age group using three statistical methods: non-parametric (Non-par), parametric on log-transformed data (Par-log), and linear regression on log-transformed data (Reg). Specificity of normal range for IGF1 was set at 95% for either lower or upper limit. For linear regression method, 90% individual confidence intervals were computed for each age group using the midpoint of each age group as age value.

### Table 2 Normal range for IGF1 using different methods. Normal range for IGF1 was evaluated in 849 healthy subjects using three different methods: non-parametric (Non-par), parametric on log-transformed data (Par-log), and linear regression on log-transformed data (Reg). Specificity of normal range for IGF1 was set at 95% for either lower or upper limit. For linear regression method, 90% individual confidence intervals were computed for each age group using the midpoint of each age group as age value.

<table>
<thead>
<tr>
<th>Age range (years)</th>
<th>n</th>
<th>Non-par Low</th>
<th>Non-par Up</th>
<th>Par-log Low</th>
<th>Par-log Up</th>
<th>Reg Low</th>
<th>Reg Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>20–24</td>
<td>78</td>
<td>178</td>
<td>544</td>
<td>151</td>
<td>405</td>
<td>175</td>
<td>504</td>
</tr>
<tr>
<td>25–29</td>
<td>85</td>
<td>151</td>
<td>407</td>
<td>138</td>
<td>370</td>
<td>133</td>
<td>383</td>
</tr>
<tr>
<td>30–34</td>
<td>73</td>
<td>138</td>
<td>379</td>
<td>123</td>
<td>340</td>
<td>120</td>
<td>345</td>
</tr>
<tr>
<td>35–39</td>
<td>79</td>
<td>123</td>
<td>324</td>
<td>108</td>
<td>299</td>
<td>110</td>
<td>314</td>
</tr>
<tr>
<td>40–44</td>
<td>99</td>
<td>101</td>
<td>289</td>
<td>92</td>
<td>246</td>
<td>82</td>
<td>237</td>
</tr>
<tr>
<td>45–49</td>
<td>82</td>
<td>101</td>
<td>289</td>
<td>92</td>
<td>246</td>
<td>82</td>
<td>237</td>
</tr>
<tr>
<td>50–54</td>
<td>72</td>
<td>95</td>
<td>272</td>
<td>76</td>
<td>239</td>
<td>88</td>
<td>251</td>
</tr>
<tr>
<td>55–59</td>
<td>89</td>
<td>76</td>
<td>256</td>
<td>66</td>
<td>215</td>
<td>70</td>
<td>202</td>
</tr>
<tr>
<td>60–64</td>
<td>69</td>
<td>66</td>
<td>202</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65–69</td>
<td>69</td>
<td>66</td>
<td>202</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70–74</td>
<td>74</td>
<td>66</td>
<td>202</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75–79</td>
<td>20</td>
<td>66</td>
<td>202</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
different statistical methods and 90% confidence limits, which correspond to a specificity of 95%. The three methods used were i) non-parametric method: 5° and 95° percentile; ii) parametric method on log-transformed IGF1; mean ± 1.645 s.d.; and iii) linear regression method on log-transformed data (both IGF1 and age). Through the linear regression equation, 90% individual CIs were computed for each age group using the midpoint of each age group as age value. The relationship between age and BMI and between BMI and log-transformed IGF1 was assessed by linear regression. The relationship between BMI and log-transformed IGF1 adjusted for age was assessed by multiple linear regression, partial r was also computed. 95% CI of a proportion was computed by Clopper–Pearson method.

The concordance between the two methods was expressed as global agreement, Kappa index, positive agreement (PA), and negative agreement (NA). The global agreement was calculated as rate between the sum of the number of subjects considered normal or abnormal from both the methods and the total number of subjects. The Kappa index is a global agreement adjusted for the chance (observed global agreement—chance global agreement)/(1—chance global agreement)). Landis and Koch have proposed the following as standards for strength of agreement for the kappa index: ≤0 = poor, 0.01–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, and 0.81–1 = almost perfect. PA and NA are the agreement relative to each rating category individually. With binary ratings, PA estimates the conditional probability, given that one of the raters, randomly selected, makes a positive rating (abnormal), the other rater will also do so. NA estimates the conditional probability, given that one of the raters, randomly selected, makes a negative rating (normal), the other rater will also do so (21). S.E.M. and 95% CI of PA and NA were computed by the Mackinnon’s formulas (22). A P value of <0.05 was considered significant. Statistical analysis was conducted with Stata version 10 and Stat-Xact version 4.

**Results**

**Identifying patients with GHD on the basis of different peak GH cut-off limits**

Proportion of patients classified as GHD greatly differed, ranging from 190 (62.9%) to 75 (24.8%) out of 302 patients, when different peak GH cut-off limits (Table 1) were used. Pituitary diseases more frequently associated with GHD were tumors of the pituitary region, including pituitary adenomas, and pituitary hypoplasia (Table 1). On the contrary, brain trauma and empty sella were less frequently associated with GHD, irrespective of peak GH cut-off limits.

**The value of a single IGF1 measurement in identifying patients with adult GHD depends on the adopted peak GH cut-off limits**

As reported in Assays section, we used the regression method for determining IGF1-normative levels owing to inconsistencies among the different age classes. Using this method, IGF1 was below the lower limit in 111 (36.6%, 95% CI, 32–42) out of 302 patients. This proportion was intermediate between that of Method 2 and Method 4 and similar to that of Method 3 (Table 1).

Among patients classified as GHD by Methods 1, 2, 3, and 4, 55.8, 62, 67.3, and 73.3% were identified as pathologic (i.e. with low IGF1 values) by a single IGF1 measurement. On the contrary, among subjects defined normal by Methods 1, 2, 3, and 4, 95.5, 91.0, 81.5, and 75.3% individuals had age-normal IGF1 concentrations respectively.

**Identifying patients with GHD on the basis of pituitary hormone deficit**

Patients were stratified keeping into consideration the absence or the presence of one, two, or three pituitary hormone deficits, except than GH (Table 3); among patients with three deficits, 95.5% were diagnosed with GHD by Method 1 (with high sensitivity) and 62.7%

![Table 3](https://via-free-access.sciencemag.org/content/164/11/5256/F1.large.jpg)

**Table 3** Identification of GH deficiency (GHD) among patients stratified on the basis of pituitary hormone deficits. Peak GH cut-off limits were as follows: Method 1: for patients with a body mass index (BMI) < 25 kg/m², a peak of GH ≤ 11.5 μg/l; for BMI 25–30 kg/m², a peak GH ≤ 8 μg/l; for BMI > 30 kg/m², a peak GH ≤ 4.2 μg/l (7). Method 2: 4.1 μg/l without BMI correction (8). Method 3: ≤ 2.5 μg/l without BMI correction (10). Method 4: for patients with a BMI < 25 kg/m², a peak of GH ≤ 2.3 μg/l; for BMI 25–30 kg/m², a peak GH ≤ 1.5 μg/l; for BMI > 30 kg/m², a peak GH ≤ 1.0 μg/l (17). No. of abnormal: number of patients with peak GH below the cut-off limit of single method; for IGF1, number of patients with IGF1 below the lower normal limits for age.

![No. with abnormal levels (%)](https://via-free-access.sciencemag.org/content/164/11/5256/F1.large.jpg)

*www.eje-online.org*
by Method 4 (with high specificity). Among patients without pituitary hormone deficit, 59.4% were recognized as normal by Method 1 and 96.8% by Method 4 respectively.

Different proportions of subjects were classified as normal or GHD when other peak GH cut-off limits were used (Table 3). Among patients with three pituitary hormone deficits, 82.1% (95% CI, 70.8–90.4) had low IGF1 values, whereas 84.5% (95% CI, 77.8–89.8) of those without pituitary hormone deficits had normal IGF1 values.

Proportion of patients classified as GHD progressively increased with the extension of hypopituitarism, regardless of the method used for the diagnosis (Table 3).

**Agreement between IGF1 and peak GH cut-off limits**

Next, we evaluated how many patients with total, partial, or absent hypopituitarism and identified as GHD or normal by single peak GH cut-off limits were similarly identified as GHD or normal by a single IGF1 measurement.

In patients with total hypopituitarism, the agreement between IGF1 and Method 1 was high (global agreement = 87%; PA = 92% and NA = 40%), being 86 and 100% for those identified as GHD or normal respectively by Method 1. On the contrary, in patients without pituitary hormone deficit, the agreement between Method 1 and IGF1 was low (global agreement = 72%; PA = 50% and NA = 81%), being 98% for subjects classifies as normal and only 35% for those identified as GHD respectively by Method 1 (Table 4).

Agreement between IGF1 and Method 4 followed an opposite way, being high in subjects without pituitary hormone deficit (global agreement = 85%; PA = 21% and NA = 92%) and low in patients with total hypopituitarism (global agreement = 69%; PA = 78% and NA = 43%). The agreement between IGF1 and Method 2 or Method 3 was intermediate in patients with total hypopituitarism (global agreement = 87 and 76% respectively) and in those without pituitary hormone deficit (global agreement = 81 and 83% respectively) (Table 4).

The agreement between IGF1 and any method in patients with partial hypopituitarism was similar and lower than that observed in patients with total hypopituitarism or those without pituitary hormone defects: global agreement was 54, 56, 62, and 60% for Methods 1, 2, 3, and 4 respectively (Table 4). The impact of specific pituitary disease on GHD is shown in Table 5.

**Patients considered GHD on the basis of IGF1 according to hypopituitarism extension**

Patients with three pituitary hormone deficits

Among the patients with total hypopituitarism (n = 67), 55 had low IGF1 levels: 100 and 98.2% were also identified as GHD by Method 1 and Method 2 respectively.

**Table 4 Agreement between different peak GH cut-off limits and IGF1 in identifying patients with GH deficiency (GHD) stratified on the extension of hypopituitarism. Figures within parentheses indicate values at 95% CI. The global agreement (GA) was calculated as rate between the sum of the number of subjects considered normal or abnormal from both methods and the total number of subjects. The Kappa index (K) is a GA adjusted for the chance; positive agreement (PA) and negative agreement (NA) are the agreement relative to each rating category individually (see Subjects and methods for details).**

<table>
<thead>
<tr>
<th>No. of deficits</th>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
<th>Method 4</th>
<th>IGF1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
<td>Abnormal</td>
<td>Normal</td>
<td>Abnormal</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>90</td>
<td>41</td>
<td>108</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>22</td>
<td>6</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td>0.72</td>
<td>0.81</td>
<td>0.83</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>0.38 (0.23–0.49)</td>
<td>0.45 (0.30–0.59)</td>
<td>0.28 (0.13–0.44)</td>
<td>0.16 (0.05–0.28)</td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>0.50 (0.38–0.63)</td>
<td>0.55 (0.41–0.70)</td>
<td>0.38 (0.19–0.57)</td>
<td>0.21 (0.01–0.40)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>0.81 (0.75–0.86)</td>
<td>0.88 (0.83–0.92)</td>
<td>0.90 (0.86–0.94)</td>
<td>0.92 (0.88–0.95)</td>
<td></td>
</tr>
<tr>
<td>1 or 2</td>
<td>14</td>
<td>34</td>
<td>19</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>29</td>
<td>6</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td>0.54</td>
<td>0.56</td>
<td>0.62</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>0.17 (0.01–0.33)</td>
<td>0.19 (0.00–0.37)</td>
<td>0.26 (0.05–0.47)</td>
<td>0.15 (0.00–0.37)</td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>0.61 (0.49–0.73)</td>
<td>0.60 (0.47–0.72)</td>
<td>0.60 (0.47–0.73)</td>
<td>0.47 (0.31–0.62)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>0.43 (0.28–0.58)</td>
<td>0.52 (0.38–0.66)</td>
<td>0.64 (0.52–0.76)</td>
<td>0.68 (0.57–0.78)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>9</td>
<td>4</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>55</td>
<td>1</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td>0.87</td>
<td>0.87</td>
<td>0.76</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>0.35 (0.17–0.54)</td>
<td>0.41 (0.20–0.62)</td>
<td>0.28 (0.05–0.52)</td>
<td>0.25 (0.04–0.46)</td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>0.92 (0.87–0.97)</td>
<td>0.92 (0.87–0.97)</td>
<td>0.85 (0.77–0.92)</td>
<td>0.78 (0.69–0.87)</td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>0.40 (0.09–0.71)</td>
<td>0.47 (0.18–0.76)</td>
<td>0.43 (0.20–0.66)</td>
<td>0.43 (0.23–0.63)</td>
<td></td>
</tr>
</tbody>
</table>
respectively and only 69% by Method 4. Among the remaining 12 patients with normal IGF1, nine (75%) and eight (66.7%) were identified as GHD by Methods 1 and 2 respectively (Table 4).

**Patients without hormone deficits** Among the patients without pituitary hormone deficit (n=155), 131 (84.5%) had normal serum IGF1 levels; Method 4 (with high specificity) identified 98.5% of these subjects as normal, whereas Method 1 (with high sensitivity) only 68.7% and Method 2 82.4%. Among the 24 patients with abnormal IGF1 levels, only three (12.5%) were considered abnormal by Method 4 also (Table 4).

**Patients with one or two pituitary hormone deficits** Of the 80 patients, 32 (40%) with partial pituitary hormone deficit had low IGF1 levels; Method 1 (with high sensitivity) identified 98.5% of these subjects as either healthy or GHD by different cut-offs. This issue has been clearly pointed out in this study: in fact, proportion of GHD patients ranged from 24.8 to 62.9% depending on the peak GH cut-off limits (Table 1). A question arising is why are cut-off limits different and how are they were obtained. Authors have used two different populations for identifying peak GH cut-off limits: normal and diseased subjects; the former were identified as healthy subjects without pituitary disease and the latter as pituitary patients with two (9) or three (17) pituitary hormone deficits; the fact that no other method is currently available to ascertain diseased subjects other than the presence of other pituitary hormone deficits carries a risk of misclassification. As a matter of fact, the range of peak GH of healthy subjects and that of GHD patients partially overlapped in some studies (17), leading to cut-off limits with divergent sensitivity and specificity. This is likely the reason why the different methods adopted worldwide and analyzed in this study differently classified the same subjects. Thus, Method 1 (and similar methods with high sensitivity) has the great potential of not missing GHD patients, however with a considerable risk of classifying a proportion of normal subjects (up to 25% of GH-sufficient individuals when Method 1 is employed (17)) with GHD; this concept was further reinforced when Method 4, the cut-off limits of which were established using the same patients and the same peak GH values utilized for defining Method 1, was employed; Method 4 showed a low sensitivity in identifying GHD among patients with three pituitary hormone deficits, whereas it classified as normal in 96.8% subjects without pituitary hormone deficits. The logical consequence of these results is that Method 1 is more suitable for identifying GHD among patients at high risk (i.e. with pituitary hormone deficits), whereas Method 4

### Table 5 Distribution of pituitary hormone deficits in each group of pituitary pathology. Other tumors refers to craniopharyngiomas, gliomas, or similar tumors of the hypothalamic–pituitary region. Others refers to miscellaneous conditions (cavernous sinus aneurism, autoimmune hypophysitis, etc.). Pituitary hypoplasia was a magnetic resonance imaging finding.

<table>
<thead>
<tr>
<th>No. of hormone deficits</th>
<th>Pituitary adenoma, n (%)</th>
<th>Other tumors, n (%)</th>
<th>Empty sella, n (%)</th>
<th>Brain trauma, n (%)</th>
<th>Pituitary hypoplasia, n (%)</th>
<th>Hydiopathic, n (%)</th>
<th>Others, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>80 (48.2)</td>
<td>7 (21.2)</td>
<td>25 (61.0)</td>
<td>24 (80.0)</td>
<td>2 (40.0)</td>
<td>3 (37.5)</td>
<td>14 (73.7)</td>
</tr>
<tr>
<td>1</td>
<td>30 (18.1)</td>
<td>1 (3.0)</td>
<td>11 (26.8)</td>
<td>4 (13.3)</td>
<td>1 (20.0)</td>
<td>3 (37.5)</td>
<td>2 (10.5)</td>
</tr>
<tr>
<td>2</td>
<td>19 (11.4)</td>
<td>6 (18.2)</td>
<td>2 (4.9)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (12.5)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>3</td>
<td>37 (22.3)</td>
<td>19 (57.6)</td>
<td>3 (7.3)</td>
<td>2 (6.7)</td>
<td>2 (40.0)</td>
<td>1 (12.5)</td>
<td>3 (15.8)</td>
</tr>
</tbody>
</table>

www.eje-online.org

Discussion

Current diagnosis of adult GHD is based on the evidence of impaired GH secretion (1, 2, 5–7) after appropriate stimulation. Although GHRH-Arg is the most used test (1, 2, 5–7, 16, 20), peak GH cut-off limits adopted by Medical Societies are different (5–7, 11, 12, 17). The practical consequence is that patients may be classified as either healthy or GHD by different cut-offs. This issue has been clearly pointed out in this study: in fact, proportion of GHD patients ranged from 24.8 to 62.9% depending on the peak GH cut-off limits (Table 1). A question arising is why are cut-off limits different and how are they were obtained. Authors have used two different populations for identifying peak GH cut-off limits: normal and diseased subjects; the former were identified as healthy subjects without pituitary disease and the latter as pituitary patients with two (9) or three (17) pituitary hormone deficits; the fact that no other method is currently available to ascertain diseased subjects other than the presence of other pituitary hormone deficits carries a risk of misclassification. As a matter of fact, the range of peak GH of healthy subjects and that of GHD patients partially overlapped in some studies (17), leading to cut-off limits with divergent sensitivity and specificity. This is likely the reason why the different methods adopted worldwide and analyzed in this study differently classified the same subjects. Thus, Method 1 (and similar methods with high sensitivity) has the great potential of not missing GHD patients, however with a considerable risk of classifying a proportion of normal subjects (up to 25% of GH-sufficient individuals when Method 1 is employed (17)) with GHD; this concept was further reinforced when Method 4, the cut-off limits of which were established using the same patients and the same peak GH values utilized for defining Method 1, was employed; Method 4 showed a low sensitivity in identifying GHD among patients with three pituitary hormone deficits, whereas it classified as normal in 96.8% subjects without pituitary hormone deficits. The logical consequence of these results is that Method 1 is more suitable for identifying GHD among patients at high risk (i.e. with pituitary hormone deficits), whereas Method 4
is for identifying normal subjects among those at low risk (i.e. without pituitary hormone deficits). Cut-off limits with high specificity, avoiding lifelong GH therapy in most healthy subjects, are likely preferable, when a single test should be used. However, high-specificity methods identified a proportion of diseased patients among those without pituitary hormone deficits, thus supporting the existence of isolated GHD, although in a small subset of patients (Table 3).

Following this view, IGF1 correctly identified 84.5% of normal subjects among those without pituitary hormone deficits and 82% GHD among those with three pituitary hormone deficits.

It is interesting to note that in order to reduce misclassification of healthy subjects to 5% (thus having 95% specificity), Corneli et al. (17) and Biller et al. (9), found similar cut-off points, although using different study populations: from ≤1 to ≤2.3 μg/l depending on BMI and ≤1.5 μg/l respectively. Unfortunately, these cut-offs have poor sensitivity (on average 70%), thus missing many GHD patients among those with total hypopituitarism. Taken together, these data suggest that the context in which the cut-offs are used is crucial, providing further evidence that the initial definition of healthy and diseased subjects does influence specificity and sensitivity of cut-off limits.

However, data from this study also throws novel light on IGF1, which has a sensitivity–specificity pair >80%, when used in an appropriate clinical context. First of all, we defined normative values for IGF1 with 95% specificity in a large series of healthy subjects; this allowed a strict definition of diseased subjects with serum IGF1 below the normal value for age. The fact that 82% of patients with total hypopituitarism had serum IGF1 below the normal range for age clearly indicates that measurement of IGF1 in this context is a good indicator of GHD, in keeping with previous reports (10, 23–27); moreover, our findings overtake any specific cut-off point and extend the concept that IGF1 below strictly defined normative range for age identified most GHD patients when applied in the appropriate clinical context. However, several points, including local specific cut-offs, should be taken into consideration for IGF1 measurements, as recently outlined (28). International guidelines for the determination of IGF1, from blood collection to data analysis, including utilization of the novel WHO IRR 02/254, will probably help achieve uniform IGF1 determinations.

IGF1 and peak GH had a high agreement in identifying GHD patients among subjects with total hypopituitarism and normal subjects among those without pituitary hormone deficits (Table 4). The highest disagreement was observed among subjects without pituitary hormone deficits (Method 1) or with total hypopituitarism (Method 4), owing to the specific feature of the methods (high sensitivity or high specificity, for Methods 1 and 4 respectively).

We suggest using IGF1 to identify normal subjects among those without pituitary hormone deficits; in this case, the remaining patients with abnormal IGF1 (14.5% of the present series) should be tested using Method 4. Likewise, IGF1 can be used for identifying patients with GHD among those with three pituitary hormone deficits; using this algorithm, the remaining
patients with normal IGF1 (17% of the present series) and high GHD risk (owing to the presence of three pituitary hormone deficits) should be submitted to a stimulatory test and evaluated using a method with high sensitivity (Method 1 or 2). Using this algorithm, only 36 (16%) patients among the 222 with three or without pituitary hormone deficits would undergo the GHRH-Arg stimulatory test (Fig. 1).

More complex is the analysis of patients with partial hypopituitarism; IGF1 can be used for identifying GHD patients (40% of the present series); the remaining patients (60% of the present series) with normal IGF1 (and one or two pituitary hormone deficits) should be evaluated using a stimulatory test; however, which cut-off limits should be adopted cannot be inferred from the present analysis; among 48 subjects, 34 (70.8%) or 29 (60.4%) were considered GHD by Method 1 or 2 respectively and only 14 (29%) by Method 4.

In conclusion, our analysis suggests that the use of single cut-off limits carries a relevant risk of misclassification of subjects screened for GHD; which cut-off limit should be used depends mainly on the clinical context; IGF1, when appropriate normative age-corrected limits are available, may be used for the first screening, avoiding a stimulatory test in most healthy or diseased subjects.

Declaration 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
This work was partially supported by Grants from the University of Pisa (Fondi d’Ateneo) to F Bogazzi and E Martino.

Acknowledgements
We thank Prof. Aldo Pinchera (University of Pisa) for his continuous encouragement and advice and the Nando Peretti Foundation (Rome, Italy) for logistic support.

References


Received 14 January 2011
Accepted 9 February 2011