Response to GH treatment in adult GH deficiency is predicted by gender, age, and IGF1 SDS but not by stimulated GH-peak

Ulla Feldt-Rasmussen1, Georg Brabant2, Dominique Maiter3, Björn Jonsson4, Andy Toogood5, Maria Koltowska-Haggstrom6, Aase Krogh Rasmussen1, Michael Buchfelder7, Bernhard Saller8 and Beverly M K Biller9

1Department of Medical Endocrinology, PE 2132, Rigshospitalet, National University Hospital, Copenhagen University, Blegdamsvej 9, DK-2100 Copenhagen, Denmark, 2Department of Endocrinology, Christie Hospital, Manchester M20 4BX, UK, 3Department of Endocrinology and Nutrition, University Hospital Saint-Luc, Brussels 1200, Belgium, 4Department of Women’s and Children’s Health, Uppsala University, Uppsala, Sweden, 5Department of Endocrinology, University Hospital Birmingham NHS Foundation Trust, Birmingham B15 2TH, UK, 6Pfizer Inc., Sollentuna, Sweden, 7Department of Neurosurgery, University of Erlangen Nuernberg, Erlangen 91045, Germany, 8Endocrine Care, Pfizer Inc., Tadworth, UK and 9Neuroendocrine Unit, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

(Correspondence should be addressed to U Feldt-Rasmussen; Email: ufeldt@rh.dk)

Abstract

Objective: We studied whether the severity of GH deficiency (GHD) defined as i) GH-peak on stimulation tests (insulin tolerance test (ITT), arginine, and glucagon), ii) number of additional pituitary deficits, or iii) baseline IGF1 SDS could impact the response to GH treatment. We further explored whether iv) IGF1 SDS after 24 months of GH replacement or v) ΔIGF1 SDS from baseline to 24 months was related to the phenotypic response to GH treatment.

Design, patients, and measurements: The patient cohort (n=1752; 50% women) was obtained from KIMS (Pfizer International Metabolic Database). The patients were divided into three groups of approximately equal size (tertiles) according to the stimulated GH-peak values and baseline IGF1 SDS and were studied at baseline, 12, and 24 months of GH therapy.

Results: Lower baseline IGF1 SDS predicted better response in weight, BMI, total cholesterol, and triglycerides, while IGF1 SDS after 24 months was associated with reduction in waist/hip ratio, total cholesterol, and improved quality of life (QoL). Age-correlated negatively with the response in body weight, BMI, waist, IGF1 SDS, and total and LDL-cholesterol.

Response in weight and BMI was greater in men than in women, whereas women showed greater improvement in QoL than men. Patients with more severe GHD as assessed by lower GH-peaks and more pituitary hormone deficiencies had a greater increase in IGF1 SDS. The increase in IGF1 SDS was associated with a reduction in waist/hip ratio and an increase in weight, BMI, and triglycerides. There was no correlation with other lipids, blood pressure, or glucose.

Conclusion: Our findings indicate that baseline and 24 months, IGF1 and its degree of increase during GH replacement were more important than stimulated peak GH to predict the phenotypic response.

European Journal of Endocrinology 168 733–743

Introduction

Severe GH deficiency (GHD) in adults causes an impairment of quality of life (QoL) (1, 2), abnormal body composition (3, 4, 5), and adverse cardiovascular risk factors (6) that may contribute to the twofold increase in mortality observed in patients with hypopituitarism and not receiving GH replacement therapy (7, 8). In children with GHD, the growth velocity serves as a biological marker, which has led to a prediction model for GH responsiveness in children (9). In adults, however, the features of GHD are not pathognomonic for the condition. Thus, endocrinologists evaluating adults have to rely upon biochemical assessment by a provocative test to identify patients with severe GHD (10). Furthermore, the response to GH replacement is assessed by surrogate markers such as phenotypic features, QoL evaluation, and measurement of insulin-like growth factor 1 (IGF1) (11).

Clinicians prescribing GH to patients with GHD wish to know which patients are most likely to respond to GH with a measurable improvement, i.e. whether the tests used for the diagnosis of GHD before treatment is started can help predict those who will experience the best clinical benefit from replacement. A few prediction models have been presented (12, 13, 14), one of which indicated that gender, body height, baseline lean body mass, and serum insulin levels were the major clinical predictors (12). It is well known that the response to GH treatment is highly individual, and it is therefore of
clinical interest to determine whether clinical variables measured at the time of GHD diagnosis and during GH replacement can suggest which patients may ultimately respond better in the long-term.

In our previous studies we showed that even hypopituitary patients with severe GHD, defined as a GH-peak ≤ 3 μg/l at diagnostic testing, demonstrated a relationship between phenotype and the degree of GH severity (15, 16). We now investigate whether the severity of GHD defined as i) GH-peak on stimulation tests (insulin tolerance test (ITT), arginine, and glucagon) (17, 18), ii) number of additional pituitary deficits (19), or iii) baseline IGF1 SDS had an impact on the response to GH replacement. We further explored whether iv) the level of IGF1 SDS after 24 months of GH replacement or v) the change (Δ) of IGF1 SDS from baseline to 24 months was related to the phenotypic response to GH treatment.

Materials and methods

KIMS (Pfizer International Metabolic Database), a long-term safety and outcome study of GH replacement therapy in adults, represents a large pharmaco-epidemiological survey initiated in 1994 (20). In this database, GH stimulation tests are reported in a standardized fashion along with baseline and follow-up data on BMI, waist circumference, waist:hip ratio, blood glucose, HbA1c, systolic and diastolic blood pressure, and QoL, measured by QoL—Assessment of Growth Hormone Deficiency in Adults score (QoL-AGHDA) (21). Serum samples were analyzed centrally for IGF1 levels and serum lipids (total cholesterol, LDL-cholesterol, HDL-cholesterol, and triglycerides) (22). The GH levels were measured using the available assay at each investigator’s institution. We previously analyzed serum GH concentrations in relation to date of entry into the database and demonstrated that there was no time-dependent trend (15). Between 1994 and October 1997, measurements of serum IGF1 were performed at Kabi Pharmacia (Stockholm, Sweden), and thereafter at Sahlgrenska University Hospital (Gothenburg, Sweden), using the following assay methods: until November 2002, RIA after acid/ethanol precipitation of IGF-binding proteins (Nichols Institute Diagnostic, San Juan Capistrano, CA, USA); until September 2006, chemiluminescence immunoassay (Nichols Advantage System); and after September 2006, Immulite 2500 (DPC Siemens, Erlangen, Germany) (23). Intra-assay, inter-assay, and total CV were < 9% in the concentration range of 125–1046 μg/l. The assay detection limit was 13.5 μg/l. Age- and gender-specific reference ranges were used to calculate an IGF1 SDS for each patient (15, 24, 25). Serum total cholesterol, LDL-cholesterol, and triglycerides were measured centrally by standardized methods (26). Serum LDL-cholesterol was calculated according to the Friedewald formula (27). Waist and hip measurements were conducted according to KIMS Guidelines (28) circulated to all participating physicians, and BMI was calculated as body weight (kg)/height (m²).

The patient cohort was obtained from KIMS, which had 13 600 adult patients with GHD as of December 2008. Patients were included in this study if they were naive (no GH treatment before entry), or semi-naive (not on GH for at least 6 months). Only patients whose diagnosis of GHD was based upon an ITT (n = 1096), arginine stimulation test (AST; n = 400), and/or glucagon stimulation test (GST; n = 256) were included. Tests were performed within 1 year before or up to 1 month after entry into KIMS with peak GH ≤ 5.0 μg/l. In patients with more than one accepted test, the highest response was used in the analyses.

Other criteria included BMI < 50 kg/m², IGF1 SDS < 2 at baseline, and GH replacement therapy for 24 months with reported visits at baseline, 12, and 24 months. These criteria yielded a cohort of 1752 patients (881 women, 50%). To determine whether treatment response could be predicted by the severity of GHD, the patients were divided before the analysis into three groups with roughly equal numbers of subjects (tertiles). This resulted in three groups with the following ranges of GH-peaks: very severe, 0–0.29 μg/l; severe, 0.3–0.8 μg/l; and moderate, 0.81–5.0 μg/l. The cut-offs were not decided arbitrarily but driven by the data in order to obtain approximately equal group sizes (Table 1). In addition, patients were also grouped according to four prespecified IGF1 SDS ranges (< −2.0, −2.0 to −1.0, −1 to 0, and 0 to 2: Table 2).

Comparisons were made between the sub-groups at 12 and 24 months of replacement with GH for the following available variables: BMI, cholesterol (total, LDL, and HDL), triglycerides, waist circumference, waist:hip ratio, blood glucose, HbA1c, systolic and diastolic blood pressure, and QoL. QoL was determined using the QoL-AGHDA questionnaire (21), where a higher score denotes a poorer QoL and a decrease in score thus indicates improvement.

The responses in IGF1 SDS and GH sensitivity at 12 months were evaluated as predictors for the subsequent phenotypic response at 24 months. GH sensitivity was calculated as the change in IGF1 concentration from baseline to 12 months divided by the administered total amount of GH in milligrams during the first 12 months of treatment. Furthermore, associations between the variables at 24 months were analyzed using bivariate correlation analyses. Additional analyses were performed to assess for gender difference and evaluate the impact of oral estrogens as well as for use of lipid and glucose-lowering drugs. Finally, data were also analyzed using a cut-off of 3 μg/l by ITT only according to GRS gold standard for diagnosing severe adult GHD (17).
Parametric tests were used, such as independent and paired samples t-tests and one-way ANOVA, in some cases with post-hoc tests using the least significant difference (LSD) correction. In correlation analyses Pearson’s method was used. When nominal data were tested, $\chi^2$ tests and Fisher’s exact tests were utilized. Additionally, because of very complex relationships and interdependence between clinical characteristics and response to GH, a path analysis was performed.

In the path analysis, indirect effects from exogenous and endogenous variables were measured by path coefficients that are equal to and can be interpreted in the same way as the standardized $\beta$ coefficients ($-1 \leq \beta \leq 1$) gained in multiple regression analyses. Direct effect from exogenous variables was measured by simple Pearsonian correlation coefficients. The total effect was in each case the sum of direct (direct arrows between boxes in Fig. 1) and indirect effects (e.g. an arrow from the variable in question via another arrow through a second variable to the response). When calculating total effects, the indirect effects of variables, path coefficients, along a path were thus multiplied and added to the direct effect. The absolute size of the total effect mirrored the relative importance of each independent (exogenous or endogenous) variable. In the path analyses, only variables with significant coefficients were used.

In the full model on the dependent variable (change in IGF1 SDS ($\Delta$IGF1 SDS) after 24 months of GH replacement), direct influences were found from age at KIMS entry, gender, and additional pituitary deficits, and the exogenous variables were proposed to influence the intermediary, endogenous variables GH-peak, and IGF1 at baseline, that in turn were supposed to influence GH dose and $\Delta$IGF1 SDS.

The standard statistical package (SPSS) for Windows, V20.0 (SPSS, Inc.) was utilized. Data are presented as mean $\pm$ S.D. Statistical significance was considered if $P < 0.05$.

### Results

#### Baseline characteristics according to GH-peak tertiles, IGF1 subgroups, and additional pituitary deficiencies

At baseline before commencement of GH replacement, there were more females (59%) in the group with GH-peak response between 0.8 and 5.0 $\mu$g/l compared with the two other groups (44 and 49% respectively) ($P<0.001$; Table 1). The gender difference was opposite for IGF1 levels; proportionally more women had an IGF1 below $-2$ SDS (56%) or between $-2$ and $-1$ SDS (50%) compared with the higher IGF1 subgroups (42 and 43%) ($P<0.001$; Table 2).
Table 2 Baseline characteristics of the GH-deficient patients grouped according to prespecified IGF1 SDS values.

<table>
<thead>
<tr>
<th>Severity of GHD: groups based on IGF1 SDS at baseline</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>( P&lt;^a )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>741</td>
<td>457</td>
<td>372</td>
<td>182</td>
<td></td>
</tr>
<tr>
<td>IGF1 SDS</td>
<td>(&lt;-2)</td>
<td>(-2) to (-1)</td>
<td>(-1) to (0)</td>
<td>(0) to (2)</td>
<td></td>
</tr>
<tr>
<td>Gender: females (%; (n))</td>
<td>56 (417)</td>
<td>50 (227)</td>
<td>43 (160)</td>
<td>42 (77)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Childhood onset (%)</td>
<td>32.1</td>
<td>7.4</td>
<td>4.6</td>
<td>2.7</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Number of additional pituitary deficiencies (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isolated GHD</td>
<td>10.0</td>
<td>10.3</td>
<td>15.3</td>
<td>15.9</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13.7</td>
<td>19.9</td>
<td>22.3</td>
<td>26.4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>16.0</td>
<td>21.7</td>
<td>21.2</td>
<td>15.9</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>39.6</td>
<td>31.9</td>
<td>30.1</td>
<td>29.1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20.7</td>
<td>16.2</td>
<td>11.0</td>
<td>12.6</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Primary etiology(^b) (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonfunctioning pituitary adenoma</td>
<td>18.5</td>
<td>38.5</td>
<td>36.6</td>
<td>33.0</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Secreting pituitary adenoma</td>
<td>21.3</td>
<td>25.6</td>
<td>28.0</td>
<td>35.7</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Other sellar tumors</td>
<td>6.2</td>
<td>4.4</td>
<td>2.2</td>
<td>5.5</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Cranioopharyngioma</td>
<td>11.5</td>
<td>7.0</td>
<td>8.6</td>
<td>6.6</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Extrasellar tumor</td>
<td>5.7</td>
<td>4.4</td>
<td>3.2</td>
<td>3.3</td>
<td>NS</td>
</tr>
<tr>
<td>Idiopathic GHD</td>
<td>19.7</td>
<td>5.0</td>
<td>5.9</td>
<td>5.5</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Malignancy treatment</td>
<td>0.7</td>
<td>1.5</td>
<td>1.9</td>
<td>0.5</td>
<td>NS</td>
</tr>
<tr>
<td>Traumatic brain injury</td>
<td>2.2</td>
<td>1.8</td>
<td>2.4</td>
<td>1.6</td>
<td>NS</td>
</tr>
<tr>
<td>Other</td>
<td>14.3</td>
<td>11.8</td>
<td>11.3</td>
<td>8.2</td>
<td>NS</td>
</tr>
<tr>
<td>Age at pituitary disease onset (years, mean (\pm) s.d.)</td>
<td>28.3(\pm) 16.7</td>
<td>39.9(\pm) 15.4</td>
<td>42.6(\pm) 14</td>
<td>44.8(\pm) 13.4</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Age at GHD diagnosis (years, mean (\pm) s.d.)</td>
<td>35.8(\pm) 17.4</td>
<td>46(\pm) 13.9</td>
<td>47.8(\pm) 12.8</td>
<td>50.2(\pm) 11.8</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Age at KIMS start (years, mean (\pm) s.d.)</td>
<td>39.8(\pm) 14.3</td>
<td>47.6(\pm) 13.4</td>
<td>49(\pm) 12.2</td>
<td>51.5(\pm) 11.2</td>
<td>(0.001)</td>
</tr>
</tbody>
</table>

\(^a\)Over-all tests were based on one-way ANOVA and \(\chi^2\). \(P\) values indicate ‘over-all’ difference.


Figure 1 Path analysis of the complex interrelations between GH dose at 24 months, Baseline variables and change in IGF1 SDS from baseline to 24 months. Numbers in the figure represent path coefficients (\(b\)). The total effect from the variables in the model was accounted for by IGF1 SDS at baseline (\(-0.67\)), age (\(-0.39\)), additional pituitary hormone deficiencies (0.29), GH-peak at stimulation test (\(-0.19\)), GH dose at 24 months (0.08), and gender (0.05).
Comparison of very severe, severe, and moderate GH-peak groups with relation to change in phenotypic features after 24 months compared with baseline before GH replacement. Significant differences between groups were only demonstrated for HDL-cholesterol (very severe vs severe, \( P < 0.009 \)) and very severe vs moderate, \( P = 0.013 \) respectively), while IGF1 differed between all the groups (\( P = 0.001 \)).

<table>
<thead>
<tr>
<th>GH-peak groups</th>
<th>Very severe</th>
<th>Severe</th>
<th>Moderate</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Body weight (kg)</td>
<td>588 (5.8)</td>
<td>534 (5.3)</td>
<td>459 (6.7)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>26.7 (5.2)</td>
<td>26.5 (5.3)</td>
<td>27.8 (7.1)</td>
</tr>
<tr>
<td>Waist (cm)</td>
<td>1.04 (1.3)</td>
<td>0.81 (1.0)</td>
<td>0.73 (0.9)</td>
</tr>
<tr>
<td>Waist/hip ratio</td>
<td>1.04 (6.3)</td>
<td>1.18 (6.7)</td>
<td>1.18 (6.7)</td>
</tr>
<tr>
<td>Diastolic BP (mm Hg)</td>
<td>0.65 (11.2)</td>
<td>0.49 (11.0)</td>
<td>0.38 (10.5)</td>
</tr>
<tr>
<td>B-glucose (mmol/l)</td>
<td>3.35 (4.5)</td>
<td>2.81 (5.9)</td>
<td>2.78 (1.7)</td>
</tr>
<tr>
<td>HbA1c (%)</td>
<td>3.72 (6.5)</td>
<td>3.72 (6.5)</td>
<td>3.72 (6.5)</td>
</tr>
<tr>
<td>Cholesterol (mmol/l)</td>
<td>4.81 (6.5)</td>
<td>4.81 (6.5)</td>
<td>4.81 (6.5)</td>
</tr>
<tr>
<td>HDL (mmol/l)</td>
<td>0.05 (0.26)</td>
<td>0.05 (0.26)</td>
<td>0.05 (0.26)</td>
</tr>
<tr>
<td>LDL (mmol/l)</td>
<td>0.32 (0.9)</td>
<td>0.32 (0.9)</td>
<td>0.32 (0.9)</td>
</tr>
<tr>
<td>Triglycerides (mmol/l)</td>
<td>3.72 (6.5)</td>
<td>3.72 (6.5)</td>
<td>3.72 (6.5)</td>
</tr>
<tr>
<td>IGF1 SDS</td>
<td>2.67 (1.7)</td>
<td>2.67 (1.7)</td>
<td>2.67 (1.7)</td>
</tr>
</tbody>
</table>

Changes calculated as 24 months – baseline. Differences between groups tested using one-way ANOVA and tests using the LSD correction.

Women were more likely than men to have no, one, or two other pituitary deficiencies (53, 58, and 55% respectively), compared with three or four additional deficiencies (43 and 49%) (\( P < 0.001 \); data not shown). Among the women, 44% (\( n = 388 \)) used estrogens (most of them orally), with a lower proportion among the patients with the highest GH-peak (Table 1). A number of patients (132, 0.5%) used glucose-lowering drugs, distributed evenly among the subgroups. However, among the 370 (17%) patients taking lipid-lowering drugs, a lower proportion was found in the group with the highest GH-peak (Table 1). There were significantly more craniopharyngiomas in the groups with the lowest GH-peak and IGF1 SDS and in the group with most additional deficiencies (Tables 1 and 2). Idiopathic GHD was more often seen in isolated GHD, and more often had lower IGF1 SDS (\( P < 0.001 \)) but similar GH-peak. The other diagnoses were largely similar among the groups. Age at diagnosis of GHD and at KIMS start was higher in the very severe group, in the 0–2 SDS IGF1 group, and in the group with fewer additional pituitary hormone deficiencies. Childhood onset GHD was more prevalent in isolated GHD and in the subgroup with lower IGF1 SDS but was similarly distributed across the GH-peak groups (Tables 1 and 2).

**GH-peak and phenotypic response to GH replacement**

Comparisons of the GH-peak groups (very severe, severe, and moderate GHD) in relation to change in phenotypic features after 24 months of GH replacement are shown in Table 3. Compared with baseline, significant differences between the three groups were demonstrated after 2 years of treatment for HDL-cholesterol, which increased slightly more on replacement (0.05 mmol/l) in the very severe group compared with the severe group (\( P < 0.05 \)), which in turn showed no change. Body weight and waist circumference decreased more in the very severe group at 12 months (data not shown), but this difference between groups was not sustained at 24 months. QoL-AGHDA scores at both 12 (data not shown) and 24 months (Table 3) improved less in the very severe group (\(-3.7 \pm 5.5\)) compared with both the severe (\(-4.8 \pm 6.0, P < 0.009\)) and the moderate group (\(-4.8 \pm 6.5, P = 0.013\)). The change in IGF1 SDS at both 12 and 24 months was more pronounced in the very severe GH-peak group and differed between all three groups, with a mean increase of \(2.7 \pm 1.7\) in the very severe, \(2.3 \pm 1.6\) in the severe, and \(1.7 \pm 1.5\) in the moderate GH-peak group respectively (\( P < 0.001 \)). The lower the GH-peak, the more pronounced was the increase in IGF1 SDS, but the less pronounced the improvement in QoL. These associations were seen in the background of the very severe group having more men than the other two groups, and patients with craniopharyngioma and nonfunctioning pituitary adenoma, while the mean age at diagnosis of the pituitary disease as well as at start in KIMS was higher (Table 1).
Thus, a low GH-peak predicted a good IGF1 response but with a poorer QoL outcome and no significant differences in any of the other phenotypic features. Reanalyses using a cut-off of 3 μg/l showed similar results to the analyses mentioned above (data not shown).

**IGF1 SDS at baseline and phenotypic response to GH replacement**

Except for a minor relationship between blood pressure, HbA1c, lipids, and additional pituitary hormone deficiencies at 24 months, only ΔIGF1 SDS at 12 and 24 months had a consistent association with IGF1 SDS at baseline (data not shown). Table 4 shows the comparison between the four IGF1 subgroups at baseline (subgroup 1 = the lowest IGF1 SDS) regarding changes in phenotypic features after 24 months. Significantly larger increases were demonstrated in the lowest IGF1 SDS subgroup for body weight (group 1 vs groups 2, 3, and 4, \( P < 0.001 \)) and BMI (group 1 vs groups 2 and 4, \( P < 0.003 \) vs group 3, \( P < 0.05 \)).

**Number of additional pituitary insufficiencies at baseline and phenotypic response to GH replacement**

There were statistically significant differences between groups characterized by the number of additional pituitary hormone deficiencies, in terms of gender, age, etiology, onset of disease, age at onset of the pituitary disease, and entrance into KIMS (data not shown). Waist circumference showed a more marked decrease in the group without additional deficiencies compared with the others (\( P < 0.04 \)), and increasing number of deficiencies resulted in a larger increment of IGF1 SDS. No other associations were identified.
Other baseline features influencing the phenotypic response to GH replacement

Patients with a high body weight at baseline decreased their weight more than leaner subjects ($r = -0.10, P < 0.0001$) and age was also negatively correlated to the response to GH in weight ($r = -0.15, P < 0.0001$), BMI ($r = -0.12, P < 0.0001$), waist circumference ($r = -0.06, P < 0.05$), total cholesterol ($r = -0.08, P < 0.01$), and LDL-cholesterol ($r = -0.09, P < 0.01$) (data not shown).

Compared with men, women decreased weight ($P < 0.05$) and BMI ($P < 0.05$) to a lesser degree but increased QoL more ($P < 0.01$). The GH-peak and number of additional hormone deficiencies predicted only the $\Delta$IGF1 SDS ($r = -0.23$ and $0.18$ respectively $P < 0.0001$; data not shown).

Prediction of GH response at 24 months by first year response to GH

The GH sensitivity in the whole cohort of patients demonstrated a bell-shaped distribution. GH sensitivity predicted the phenotypic response from 12 to 24 months in diastolic blood pressure ($r = -0.09, P < 0.001$), total cholesterol ($r = -0.07, P < 0.05$), and IGF1 SDS ($r = -0.31, P < 0.0001$). IGF1 SDS at 12 months predicted the response in waist circumference ($r = 0.07, P < 0.05$) and total cholesterol ($r = -0.06, P < 0.05$), while change in IGF1 SDS within the first year predicted the change of BMI ($r = 0.06, P < 0.05$), blood glucose ($r = -0.10, P < 0.01$), and the change of IGF1 SDS ($r = -0.31, P < 0.0001$). The negative association between peak GH and baseline and $\Delta$IGF1, respectively, was slightly attenuated by controlling for estrogen use, but still highly significant ($P < 0.00001$; data not shown).

Association between outcomes of 24 months’ GH replacement

$\Delta$IGF1 SDS correlated positively with the change in weight ($r = 0.13, P < 0.0001$; Fig. 2, Table 5), BMI ($r = 0.11, P < 0.0001$), and triglycerides ($r = 0.06, P < 0.05$), and negatively with waist/hip ratio ($r = -0.07, P < 0.05$). In other words, an increase in IGF1 SDS after 24 months was associated with a reduction in waist/hip ratio. IGF1 SDS at 24 months was negatively correlated with a change in waist/hip ratio ($r = -0.06, P < 0.05$), change in QoL-AGHDA score ($r = -0.09, P < 0.01$; in men NS; in women: $r = -0.17, P < 0.001$), and change in total cholesterol ($r = -0.06, P < 0.05$), the latter of which disappeared when controlling for use of lipid lowering drugs.

Path analysis of the change of IGF1 SDS from baseline to 24 months, $\Delta$IGF1 SDS

Path analysis confirmed the complexity of interrelations between the variables involved in the change of IGF1 SDS in response to GH (Fig. 1, Table 5). IGF1 SDS at baseline had the most important influence on $\Delta$IGF1 SDS ($\beta = -0.64$) and this influence was further strengthened to $-0.67$, through the indirect effect via

---

**Figure 2** Scattergrams showing change ($\Delta$) in (A) body weight and (B) waist hip ratio vs change in IGF1 SDS after 24 months of GH replacement therapy of adult patients with GHD and change (C) in body weight vs age at KIMS entry.
GH dose. Thus, patients with a low-basal IGF1 SDS had a greater ΔIGF1 SDS at 24 months than patients with higher IGF1 SDS levels.

Age at KIMS start had a direct influence on ΔIGF1 SDS (r = 0.12). This effect was strengthened by the paths via IGF1 SDS at baseline and GH dose and weakened somewhat by the number of additional deficits (r = 0.06). The total effect of age was −0.41, i.e. with increasing age, IGF1 SDS changed less under GH replacement therapy.

The number of pituitary hormone deficits in addition to GHD had a direct effect on ΔIGF1 SDS (r = 0.18). The total effect was 0.32, reached through the paths via all the intermediary predictors.

Gender by itself did not have a significant correlation with ΔIGF1 SDS. The total influence of gender was thus weak (0.05) and reached through paths via all intermediary predictors but weakened by the reciprocal correlation with the number of additional pituitary deficits.

The GH-peak had an indirect path to ΔIGF1 SDS with a low β coefficient (−0.04), but the total effect was enforced by another indirect effect via basal IGF1 SDS and therefore amounted to −0.19.

GH dose at 24 months was at the last position in the chain of predictors, and thus had only one path to ΔIGF1 SDS with a weak β coefficient of 0.08 (Fig. 1).

### Discussion

The response to GH replacement in adults is less predictable than in children and includes many...
variables such as changes in body composition, QoL, and blood lipids (30, 31), and many confounders have to be taken into account (32, 33). Prediction models in children with short stature incorporated data from a very large number of patients in the calculations (9, 34, 35). In contrast, prediction of GH response has been described in only a limited number of adults with GHD (12, 13, 14), and we therefore explored predictors of response using the large KIMS database (20).

The original and follow-up consensus statements of the Growth Hormone Research Society identified ITT as a gold standard stimulation test for diagnosing severe GHD (peak GH < 3 μg/l) in adults (17, 18). Toogood et al. (19) demonstrated that the severity of hypopituitarism as indicated by the number of anterior pituitary hormone deficits was associated with the GH response to ITT. This observation was substantiated in our previous studies of larger patient cohorts with severe GHD based on peak GH of < 3 μg/l during ITT, AST, or GST stimulation tests (15, 16). In our selection of the patient cohort for the current study, we included those who had a GHD diagnosis confirmed by any of the three tests, allowing a cut-off < 5 μg/l in order not to exclude young adult patients, whose cut-off levels are considered to be higher than in older adults (36). The cut-off of < 5 μg/l has furthermore been demonstrated to provide a more optimal sensitivity and precision in receiver operating curves (ROC) curves distinguishing hypopituitary patients from controls (37). Nevertheless, we performed analyses using the classical cut-off of < 3 μg/l which confirmed our results.

In a prior study on patients with severe GHD, the age at which the patient was tested did not affect the performance of the ITT, AST, and GST (16). This suggested that the impact of pituitary disease overrode the subtle changes that occur in the regulation of GH secretion with aging and need not be taken into account when diagnosing GHD in this clinical scenario (16, 38). On the other hand, in the current study on hypopituitary patients with GHD, older age negatively predicted the response to GH in weight, BMI, waist circumference, and total and LDL-cholesterol, in keeping with a previous description of GH replacement in the elderly (39), while others have identified elderly GHD patients as a GH-sensitive group (40, 41).

Obesity has consistently been shown to inhibit spontaneous GH secretion and all dynamic function tests that have been studied – the greater the BMI, the less GH produced (16, 33, 37, 42). Concerns have therefore been raised that obesity may compromise a reliable diagnosis of GHD because the GH response to stimulation tests will be attenuated. It is also possible that the benefit of GH replacement might be hampered by obesity. The current study was therefore reassuring that patients who were obese at baseline decreased their weight more than patients with lower body weight. Conversely, body weight and BMI both increased with increasing increment of IGF1 SDS. However, waist circumference and waist/hip ratio decreased with increasing increment of IGF1 SDS, suggesting that GH replacement therapy increased muscle mass rather than fat mass. Another possibility may be a weight gain due to water and sodium retention through renal effects of GH (43). A more precise evaluation of body composition was, however, unavailable in KIMS to test this hypothesis.

Serum IGF1 has a limited role in the diagnosis of GHD in adults (10), although the levels are highly influenced by GH status and there are correlations between the two (15). In our current study, IGF1 SDS was a strong predictor of the response to GH replacement, in agreement with some previous reports (44), but in disagreement with others (12, 13, 14). We also found that women using estrogens had lower IGF1 levels at baseline but responded better in terms of ΔIGF1. A recent study developed accurate mathematical models to predict GH responsiveness in GHD adults and found gender, body height, baseline lean body mass, and serum insulin levels as the major clinical predictors (12). We were not able to test this model, since measurements of lean body mass and insulin were not available in the KIMS database.

There are a number of limitations to our study. Unlike IGF1, GH was measured in each individual laboratory, which was a possible confounder of the results of GH-peaks. GH measurements are generally not very well standardized between methods and laboratories (45), and the period for entrance of patients into the KIMS database spanned more than 15 years. However, we have previously detected no variation in the GH-peaks over time (15). Furthermore, the size of the study should theoretically eliminate or at least diminish the influence from confounders and method variation, but it cannot be excluded that the results were partly affected by differences in GH measurements. Although central IGF1 measurements is one of the strengths of our study, it is worth mentioning that there were a number of changes in the assay methods used, which may also have influenced the results. Finally, adjustment for changes in other hormone replacements was not performed because of variability in the accuracy of concomitant medication reporting.

In summary, the extent of hypopituitarism indicated by the GH stimulation test peak and the number of additional hormone deficiencies did not have an independent impact on the response to GH but predicted only the IGF1 SDS response. IGF1 at baseline predicted the GH treatment response in weight, BMI, cholesterol, and triglycerides; age predicted the response in body weight, BMI, waist, and total and LDL-cholesterol, while gender predicted the responsiveness in weight, BMI, and QoL-AGHDA. In addition, the IGF1 response to GH treatment from baseline to 24 months, as well as the IGF1 SDS with 24 months of replacement, was associated with a positive treatment effect on weight, BMI, waist/hip ratio, and triglycerides, but not
on other lipids, blood pressure, or blood glucose, and with reduction in waist/hip, total cholesterol, and QoL-AGHDA respectively.

In conclusion, our findings have indicated that IGF1 at baseline and at 24 months, as well as its change during treatment, played a more important role than peak GH to stimulation testing in predicting the phenotypic response to GH replacement. However, the clinical implications of these observations remain to be proven in future studies.

Declaration of interest
G Brabant, B M K Biller, M Buchfelder, U Feldt-Rasmussen, D Maiter, A Toogood, and B Jonsson received consulting fees from Pfizer, Inc. U Feldt-Rasmussen received research funding from Novo Nordisk. B M K Biller received consulting fees from Novo Nordisk and research funding from Novo Nordisk, Pfizer, and Merck Serono. M Kolotowska-Haggstrom is an employee of Pfizer, Inc. (Sweden), and B Saller is an employee of Pfizer, Inc. (UK).

Funding
Pfizer sponsored the KIMS database. U Feldt-Rasmussen received a research grant from Arvid Nilsson’s Fund. Salary for her research time is paid by this fund. No authors were paid for their contributions to this manuscript.

References
Severe GHD and treatment response 743


