

# Sensitivity and Specificity of Six Tests for the Diagnosis of Adult GH Deficiency

BEVERLY M. K. BILLER, MARY H. SAMUELS, ANTHONY ZAGAR, DAVID M. COOK, BAHA M. ARAFAH, VIVIEN BONERT, STAVROS STAVROU, DAVID L. KLEINBERG, JOHN J. CHIPMAN, AND MARK L. HARTMAN

Massachusetts General Hospital (B.M.K.B.), Boston, Massachusetts 02114; Oregon Health Sciences University (M.H.S., D.M.C.), Portland, Oregon 97201; Lilly Research Laboratories (A.Z., J.J.C., M.L.H.), Eli Lilly & Co., Indianapolis, Indiana 46285; University Hospital of Cleveland (B.M.A.), Cleveland, Ohio 44106; Cedars Sinai Medical Center (V.B.), Los Angeles, California 90048; and New York University Medical Center (S.S., D.L.K.), New York, New York 10010

Although the use of the insulin tolerance test (ITT) for the diagnosis of adult GH deficiency is well established, diagnostic peak GH cut-points for other commonly used GH stimulation tests are less clearly established. Despite that fact, the majority of patients in the United States who are evaluated for GH deficiency do not undergo insulin tolerance testing. The aim of this study was to evaluate the relative utility of six different methods of testing for adult GH deficiency currently used in practice in the United States and to develop diagnostic cut-points for each of these tests. Thirty-nine patients (26 male, 13 female) with adult-onset hypothalamic-pituitary disease and multiple pituitary hormone deficiencies were studied in comparison with age-, sex-, estrogen status-, and body mass index-matched control subjects (n = 34; 20 male, 14 female). A third group of patients (n = 21) with adult-onset hypothalamic-pituitary disease and no more than one additional pituitary hormone deficiency was also studied. The primary end-point was peak serum GH response to five GH stimulation tests administered in random order at five separate visits: ITT, arginine (ARG), levodopa (L-DOPA), ARG plus L-DOPA, and ARG plus GHRH. Serum IGF-I concentrations were also measured on two occasions. For purposes of analysis, patients with multiple pituitary hormone deficiencies were assumed to be GH deficient. Three diagnostic cut-points were calculated for each test to provide optimal separation of multiple pituitary hormone deficient and control subjects according to three criteria: 1) to minimize misclassification of control subjects and deficient patients (balance between high

sensitivity and high specificity); 2) to provide 95% sensitivity for GH deficiency; and 3) to provide 95% specificity for GH deficiency. The greatest diagnostic accuracy occurred with the ITT and the ARG plus GHRH test, although patients preferred the latter (P = 0.001). Using peak serum GH cut-points of 5.1  $\mu\text{g/liter}$  for the ITT and 4.1  $\mu\text{g/liter}$  for the ARG plus GHRH test, high sensitivity (96 and 95%, respectively) and specificity (92 and 91%, respectively) for GH deficiency were achieved. To obtain 95% specificity, the peak serum GH cut-points were lower at 3.3  $\mu\text{g/liter}$  and 1.5  $\mu\text{g/liter}$  for the ITT and ARG plus GHRH test, respectively. There was substantial overlap between patients and control subjects for the ARG plus L-DOPA, ARG, and L-DOPA tests, but test-specific cut-points could be defined for all three tests to provide 95% sensitivity for GH deficiency (peak GH cut-points: 1.5, 1.4 and 0.64  $\mu\text{g/liter}$ , respectively). However, 95% specificity could be achieved with the ARG plus L-DOPA and ARG tests only with very low peak GH cut-points (0.25 and 0.21  $\mu\text{g/liter}$ , respectively) and not at all with the L-DOPA test. Although serum IGF-I levels provided less diagnostic discrimination than all five GH stimulation tests, a value below 77.2  $\mu\text{g/liter}$  was 95% specific for GH deficiency. In conclusion, the diagnosis of adult GH deficiency can be made without performing an ITT, provided that test-specific cut-points are used. The ARG plus GHRH test represents an excellent alternative to the ITT for the diagnosis of GH deficiency in adults. (*J Clin Endocrinol Metab* 87: 2067–2079, 2002)

**G**H DEFICIENCY (GHD) in adults is now a well-recognized condition, associated with a number of metabolic abnormalities, many of which are reversible with replacement therapy (1). Current published consensus guidelines recommend that the diagnosis of adult GHD be established in patients with an appropriate clinical history by demonstrating a peak GH concentration of less than 3–5  $\mu\text{g/liter}$  following insulin-induced hypoglycemia (insulin tolerance test, ITT) (2, 3). However, this test is labor intensive, has potential risks, and is contraindicated in some patients.

Abbreviations: ARG, Arginine; BMI, body mass index; CART, classification and regression tree analysis; FN, false negative; FP, false positive; GHD, GH deficiency; GHRP, GH-releasing peptides; ITT, insulin tolerance test; L-DOPA, levodopa; MPH, multiple pituitary hormone deficiencies; PHD, pituitary hormone deficiencies; PPV, positive predictive value; ROC, receiver-operating characteristic; SDS, SD scores; TN, true negative; TP, true positive.

For these reasons, the ITT is not commonly performed in the United States. A recent study of over 800 patients undergoing evaluation for GHD in the United States reported that 77.7% of the subjects were tested with arginine (ARG), levodopa (L-DOPA), or ARG plus L-DOPA, with only 11.4% of patients undergoing the ITT (4). These other commonly employed GH stimulation tests have been interpreted using the same diagnostic cut-points as those reported for the ITT (3–5  $\mu\text{g/liter}$ ), despite lack of data supporting this approach. Although the combined ARG plus GHRH test has been proposed to be the best alternative to the ITT with a peak GH cut-point of 9  $\mu\text{g/liter}$  (3, 5), this test has been reported to be performed in fewer than 1% of patients in the United States (4). While the response to multiple pharmacologic agents in healthy adults has been evaluated in several studies (6–12), comparison of diagnostic stimulation tests in patients with hypothalamic-pituitary disease has been limited to two or

three tests in prior studies (5, 13–16). Drawbacks of earlier studies included a lack of control subjects or matching control subjects to hypopituitary patients by age and sex alone. While such reports are valuable, GH secretion is also influenced by other factors, such as body composition and estrogen use (17). Therefore, evaluation of tests designed to diagnose GHD should control for such variables (18, 19).

We investigated the utility of five different stimulation tests used in clinical practice in the United States and developed test-specific cut-points to improve the diagnostic accuracy of these tests. Toward this end, we compared the peak serum GH response in patients with adult-onset hypothalamic-pituitary disease and multiple (two or more) additional (other than GHD) pituitary hormone deficiencies (MPHD) with that in control subjects rigorously matched for age, sex, body mass index (BMI), and estrogen status. The diagnostic usefulness of the GH-dependent biochemical marker, IGF-I was also evaluated. To define diagnostic cut-points for each test, we assumed that the MPHD patients were GH deficient, based on prior studies demonstrating that such patients have an approximately 90% chance of having severe GHD with the ITT (20, 21). Three diagnostic cut-points were calculated for each test, using two distinct statistical methods, to provide optimal separation of MPHD and control subjects according to three criteria: 1) to minimize misclassification of control subjects and MPHD patients (balance between high sensitivity and high specificity); 2) to provide 95% sensitivity for GHD; and 3) to provide 95% specificity for GHD. These cut-points were then used to evaluate a third group of subjects with hypothalamic-pituitary disease and not more than one additional pituitary hormone deficiency [0–1 pituitary hormone deficiencies (PHD)]. Previous studies have shown that such patients have a lower probability of having severe GHD (20, 21), but few data comparing different tests are available for this patient group.

## Subjects and Methods

### Subjects

Study subjects were recruited at five United States pituitary centers. The institutional review board at each site approved the study, and all patients gave written informed consent. Subject characteristics are shown in Table 1. The primary study group consisted of 39 patients with adult-onset hypothalamic-pituitary disease and MPHD. The four PHD considered in this study were: 1) TSH deficiency; 2) ACTH deficiency; 3) gonadotropin deficiency (LH and/or FSH deficiency were counted as one deficiency); and 4) AVP deficiency (central diabetes insipidus). PRL deficiency was not considered a PHD in this study. Women under the age of 50 yr with untreated hypogonadism were excluded from the study. The MPHD group included 26 men and 13 women; ages ranged from 26.3 to 70.1 yr (mean:  $48.9 \pm 11.1$  [sd] yr, median: 49.2 yr). The most common disorders leading to hypopituitarism in this group were pituitary adenomas (74%) or other tumors (18%), as shown in Table 2.

**TABLE 1.** Subject characteristics

	Patients with MPHD n = 39	Control subjects n = 34	Patients with 0–1 PHD n = 21
Age (yr)	48.9 ± 11.1	47.2 ± 11.3	48.2 ± 11.3
BMI (kg/m <sup>2</sup> )	30.5 ± 6.1	30.3 ± 5.8	29.2 ± 8.3
Men	26 (67%)	20 (59%)	4 (10%) <sup>a</sup>
Women	13 (33%)	14 (41%)	17 (81%) <sup>a</sup>

<sup>a</sup>  $P < 0.01$  vs. control subjects and MPHD patients.

**TABLE 2.** Hypothalamic-pituitary disease characteristics

	Patients with MPHD n = 39	Patients with 0–1 PHD n = 21
Pituitary adenomas	29 (74%)	15 (71%)
Craniopharyngiomas	6 (15%)	0
Head trauma or Sheehan's syndrome	2 (5%)	1 (5%)
Empty sella	0	2 (10%)
Sellar cyst or inflammation	0	2 (10%)
Medulloblastoma	1 (3%)	0
Surgical hypophysectomy	1 (3%)	0
Idiopathic	0	1 (5%)
Central hypogonadism	38 <sup>a</sup> (97%)	7 <sup>b</sup> (33%)
Central hypothyroidism	35 (90%)	3 (14%)
Central hypoadrenalism	31 (79%)	1 (5%)
Central diabetes insipidus	7 (18%)	1 (5%)

<sup>a</sup> All 38 patients with MPHD who had 2° hypogonadism received gonadal steroid replacement. One patient who did not have 2° hypogonadism received estrogen for treatment of menopause.

<sup>b</sup> Of the 7 patients with 0–1 PHD who had 2° hypogonadism, 4 received gonadal steroid replacement. Among the 14 patients with 0–1 PHD who did not have 2° hypogonadism, 4 women received estrogen either for contraception or treatment of menopause.

Thirty-one percent of patients had two additional (other than GHD) PHD, 54% had three PHD, and 15% had four PHD (Table 2).

A second group of 21 patients (4 men, 17 women) with adult-onset hypothalamic-pituitary disease and 0–1 PHD was also studied. The age range in these subjects was 26.5 to 65.4 yr (mean:  $48.2 \pm 11.3$  yr, median: 49.2 yr), and 71% had a history of pituitary adenoma. Forty-three percent of this group had no additional PHD; 57% had one treated PHD (Table 2).

Intracranial lesions had been stable for at least 2 yr before study entry, and at least 3 months of stable treatment were required for those taking hormone replacement for hormone deficiencies other than GHD. GH therapy was not administered for at least one month before study entry, based on data that demonstrate that serum IGF-I levels return to near-baseline within 48 h after cessation of GH (22). The protocol required that all patients with hypogonadism be treated with sex steroid replacement therapy, except for women over 50 yr of age. Patients undergoing concurrent therapy with monoamine oxidase inhibitors or cabergoline were excluded. Patients taking other dopamine agonists were required to discontinue therapy 7 d (pergolide) and 4 d (bromocriptine) before each stimulation test. All subjects (both pituitary patients and control subjects) with a history of acromegaly, active Cushing's disease, cardiovascular or cerebrovascular disease, seizures, diabetes, malignancy, renal, or hepatic dysfunction or who were pregnant were excluded from the study.

The study also enrolled 34 control subjects (20 men, 14 women), matched to the MPHD patients for sex, age ( $\pm 5$  yr), BMI ( $\pm 2$  kg/m<sup>2</sup>), and estrogen status. Their ages ranged from 24.1–68.1 yr (mean:  $47.2 \pm 11.3$  yr, median: 48.0 yr). For matching of estrogen status, women under the age of 50 yr with MPHD who were on estrogen replacement therapy were matched to female control subjects who were also taking estrogen (as an oral contraceptive or for replacement). Women with MPHD over the age of 50 yr who had untreated hypogonadism were matched to female control subjects who were not receiving estrogen. The control subjects were healthy, and had undergone normal growth and development. The female control subjects had a history of regular, age-appropriate menses. The male control subjects had normal serum testosterone concentrations. Serum PRL concentrations were normal in all control subjects.

### Study procedures

All subjects were first evaluated at a screening visit for a complete medical history, physical examination and laboratory tests. They then underwent stimulation testing on five separate mornings; each visit was separated by 5–21 d. Subjects fasted overnight for at least 10 h before

arrival and refrained from strenuous exercise on the morning of each test. The indwelling catheter was inserted 30–60 min before the initial baseline blood sampling. Five GH stimulation tests were performed in random order, one at each study visit, according to the following procedures.

1) *ITT*. Regular human insulin 0.10–0.15 U/kg was administered iv with a target blood glucose less than 40 mg/dl. Additional insulin boluses were administered if needed to achieve the target glucose value unless the investigator believed this to be unsafe. Administration of iv dextrose was allowed if the subject developed signs of neuroglycopenia in association with hypoglycemia.

2) *ARG test*. Thirty grams of L-ARG hydrochloride (10% solution) were infused iv over 30 min.

3) *L-DOPA*. L-DOPA (500 mg) was administered PO.

4) *Combined ARG plus L-DOPA test (ARG-L-DOPA)*. L-DOPA (500 mg) was given PO at initiation of the 30-min L-ARG (30 g) iv infusion.

5) *Combined ARG plus GHRH test (ARG-GHRH)*. GHRH (Geref Diagnostic, provided by Serono, Norwell, MA) 1  $\mu$ g/kg was administered by an iv bolus, followed by a 30-min infusion of L-ARG (30 g).

All tests were performed in all subjects, except for the ITT, which was omitted in subjects older than 55 yr (26 of the 94 study subjects). A twelve-lead electrocardiogram was performed before each ITT to exclude active ischemia. Blood was sampled every 20–30 min for 2.5 h, beginning 30 min before the administration of the provocative agents. During each test, blood samples were centrifuged and frozen ( $-20^{\circ}\text{C}$ ). Following completion of all stimulation tests for an individual subject, the frozen samples were shipped in a single batch on ice to a central laboratory. At the conclusion of the last stimulation test, patients were asked to rank the tests in order of preference from one (most preferred) to five (least preferred).

Serum IGF-I concentrations were measured at the screening visit and the first stimulation test visit. *SD* scores (SDS) were calculated for all patients with MPHD and for all control subjects, using the serum IGF-I means and *SD*s, appropriate for age and sex, provided by the central laboratory (Esoterix Endocrinology, Calabasas Hills, CA).

## Assays

Samples were analyzed in duplicate at a central laboratory (Esoterix Endocrinology, Calabasas Hills, CA) according to the following procedures.

*GH assay*. All serum samples from an individual subject were analyzed in one assay. Serum GH concentrations were measured using an immunochemiluminometric assay specific for 22-kDa human GH with a sensitivity of 0.05  $\mu$ g/liter (23). The intra and interassay coefficients of variation ranges were 3.8–9.1% and 8–10%, respectively for a quality control range of 0.3–20  $\mu$ g/liter. Samples higher than 20  $\mu$ g/liter were repeated on dilution. This assay is calibrated against the WHO 80/505 international GH standard (human pituitary derived GH) but uses native-sequence recombinant human GH as standard (Eli Lilly & Co., Indianapolis, IN). This method yields results that are on average one half of those obtained with a polyclonal RIA (Mark Stene, Esoterix Endocrinology, personal communication).

*IGF-I assay*. Serum IGF-I concentrations were measured in a highly specific competitive binding RIA after acid-ethanol extraction (24). The assay uses native sequence IGF-I (Bachem, Torrance, CA) as standard but is standardized 16% higher than native sequence (mass correct) IGF-I (Genentech, Inc., South San Francisco, CA) because the normal ranges were established before the availability of this standard. IGF-II is added to each assay tube to eliminate potential interference from residual low molecular weight IGF binding proteins. The assay sensitivity was 12.9  $\mu$ g/liter. The intra and interassay coefficients of variation ranges were 4.1–6.5% and 6.6–8.4% for a quality control range of 60  $\mu$ g/liter–500  $\mu$ g/liter, respectively (Mark Stene, Esoterix Endocrinology, personal communication).

## Statistical analysis

Data are presented throughout as mean  $\pm$  *SD*, except where noted otherwise. Serum GH values below the detection limit of less than 0.05  $\mu$ g/liter were set at 0.025  $\mu$ g/liter for analysis. The peak serum GH response was used as the primary variable for analysis of stimulation tests. Group differences were analyzed by *t* tests or  $\chi^2$  tests. Significance (two-sided) was set at 0.05. Patient test preferences were analyzed by Friedman's test and sign tests. The agreement between the two serum IGF-I measurements was assessed by Pearson's correlation between the two values.

To define diagnostic cut-points for each test, we assumed that the MPHD patients were GH deficient, based on prior studies demonstrating that such patients have an approximately 90% chance of having severe GHD based upon testing with the ITT (20, 21). Three diagnostic cut-points were calculated for each test, using two distinct statistical methods (described below), to provide optimal separation of MPHD and matched control subjects according to three criteria: 1) to minimize misclassification of control subjects and MPHD patients (balance between high sensitivity and high specificity); 2) to provide 95% sensitivity for GHD; and 3) to provide 95% specificity for GHD. Positive predictive value (PPV), sensitivity and specificity were calculated for each test using the numbers of patients with true positive (TP), true negative (TN), false positive (FP) and false negative (FN) results (25). In this analysis, MPHD patients were classified as TP or FN, depending on whether their peak GH or IGF-I value was below or above the test-specific cut-point, respectively. Control subjects were classified as TN or FP, depending on whether their peak GH or IGF-I value was above or below the test-specific cut-point, respectively. Sensitivity was defined as the percentage of patients with MPHD who had a peak GH below the test-specific cut-point (calculated as TP/[TP + FN]). Specificity was defined as the percentage of control subjects with peak GH above the test-specific cut-point (calculated as TN/[TN + FP]). Positive predictive value was defined as the likelihood that a subject with a positive test (peak GH below the test-specific cut-point) was clinically GH-deficient, based on the presence of MPHD (calculated as TP/[TP + FP]). After defining test-specific cut-points based on the comparison of MPHD patients and matched control subjects, the percentage of 0–1 PHD patients with peak GH values below these cut-points was calculated for each test.

Classification and regression tree analysis (CART) was performed (S-Plus 2000) to discriminate MPHD patients from the matched control subjects based upon the peak GH or IGF-I concentration, and upon IGF-I SDS. This computer algorithm calculated cut-point values for peak serum GH and IGF-I that minimized misclassification of patients with MPHD and control subjects. Diagnostic cut-points defined by CART provide a balance between high sensitivity and high specificity. The impact of age, sex, and BMI on cut-points calculated by CART was explored by including these factors in the model. Additionally, multiple linear regression analysis was used to describe the effects of age, sex and BMI on peak GH values.

The diagnostic accuracy of each test was also investigated using receiver-operating characteristic (ROC) curves calculated with AccuROC (version 2.0) (26). Once again, the comparison of MPHD patients and matched control subjects was used for this analysis, with PPV, sensitivity and specificity as defined above. The ROC curves plotted the TP rate (sensitivity) against the FP rate (1-specificity) for different cut-points of peak GH and serum IGF-I concentrations, and for IGF-I SDS. A test with perfect discrimination between the control and MPHD groups (100% sensitivity and 100% specificity) would coincide with the upper left corner of the box, and be associated with a ROC area of 1.0 (see Fig. 3). In contrast, a test providing no discrimination between the two groups would result in a diagonal line from the lower left to the upper right corner of the box (sensitivity = 1-specificity), and correspond to a ROC area of 0.5. Peak GH and serum IGF-I cut-points corresponding to 95% sensitivity and 95% specificity for GHD were calculated using the ROC curves.

Adverse events that occurred in more than 5% of subjects within 48 h of each GH stimulation test are reported for all three groups of subjects combined because they were similar in all groups. There were no serious adverse events (defined by regulatory criteria) during this study.

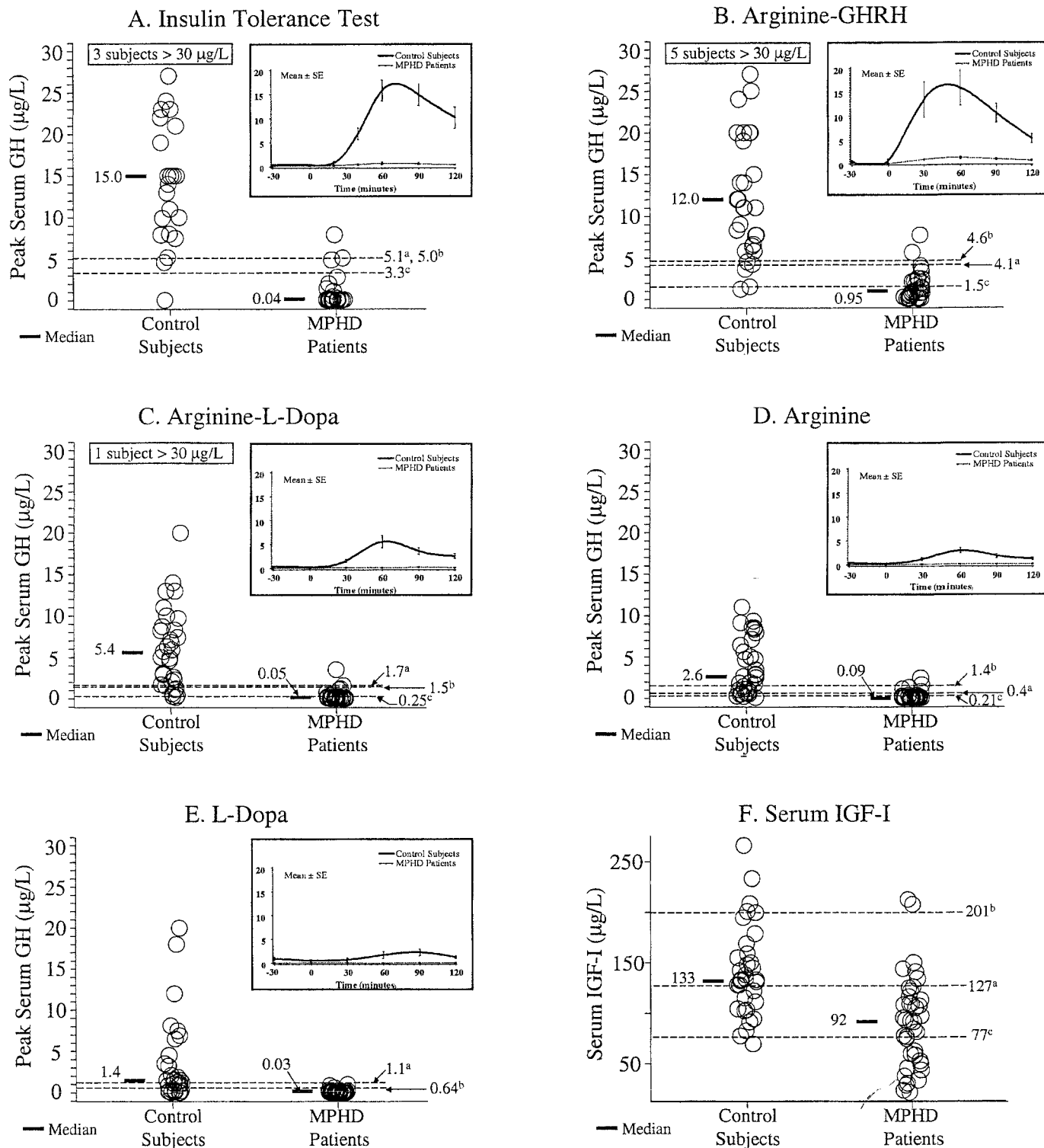


FIG. 1. Individual peak serum GH concentrations in control subjects and in patients with MPHD for: A, ITT; B, ARG-GHRH; C, ARG-L-DOPA; D, ARG; and E, L-DOPA. Panel F depicts serum IGF-I concentrations measured at the screening visit in the two groups. The median peak GH or IGF-I level in each group is denoted with a bar. The upper right box in each stimulation test panel shows the time course of the mean GH response in the two groups. The dashed lines and superscripts indicate diagnostic cut-points that can be employed (see Table 3) as follows: a, minimize misclassification of MPHD patients and control subjects; b, 95% sensitivity for GHD; and c, 95% specificity for GHD.

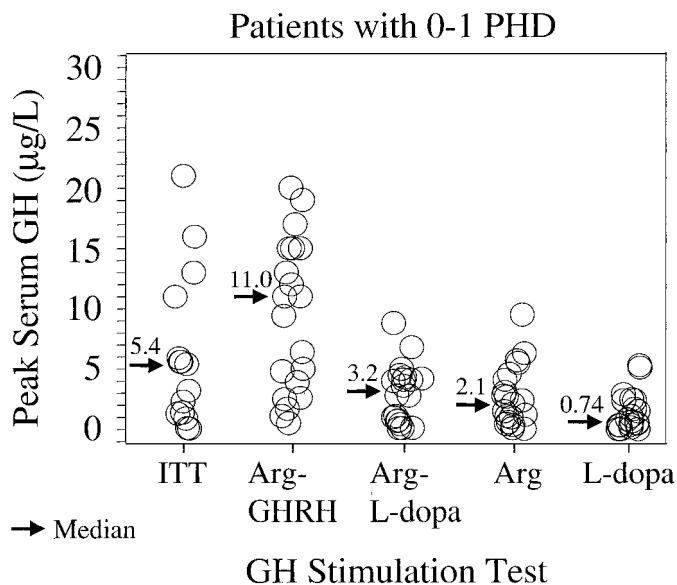


FIG. 2. Individual peak serum GH responses to each of the stimulation tests in patients with 0–1 PHD. The median for each test is denoted with an arrow.

### Results

The individual peak serum GH concentrations and the mean serum GH concentrations over time for each test in the MPHD patients and matched control subjects are shown in Fig. 1. Figure 2 shows individual peak GH concentrations in patients with 0–1 PHD. The ROC curves for each test are shown in Fig. 3. The ROC areas and test-specific cut-points determined by CART and ROC analyses (with PPV, sensitivity and specificity for each cut-point) are shown in Table 3.

#### ITT

The mean peak GH response to ITT in the patients with MPHD was  $0.95 \pm 1.9$   $\mu\text{g}/\text{liter}$  (range: 0.025–7.9  $\mu\text{g}/\text{liter}$ , median: 0.04  $\mu\text{g}/\text{liter}$ ). Ninety-three percent of these subjects had peak serum GH levels below 5  $\mu\text{g}/\text{liter}$  and 89% had levels below 3  $\mu\text{g}/\text{liter}$ .

The peak GH response in the age, BMI, sex, and estrogen-matched control subjects was significantly ( $P < 0.001$ ) higher than in the MPHD patients (mean:  $17.8 \pm 12.5$   $\mu\text{g}/\text{liter}$ ; range: 0.025–52.0  $\mu\text{g}/\text{liter}$ ; median: 15.0  $\mu\text{g}/\text{liter}$ ). One control subject (age: 38.2 yr, BMI: 35.3  $\text{kg}/\text{m}^2$ ), who had a blood glucose nadir of 45 mg/dl, had a peak GH response of less than 0.05  $\mu\text{g}/\text{liter}$  during the ITT. His peak GH response to ARG-GHRH administration was 1.5  $\mu\text{g}/\text{liter}$ , but his response to ARG-L-DOPA was above the median for the control group at 5.9  $\mu\text{g}/\text{liter}$ . All other control subjects had peak serum GH concentrations of at least 4.6  $\mu\text{g}/\text{liter}$  with ITT.

In the patients with a history of pituitary disease and 0–1 PHD, the peak GH response (mean:  $6.2 \pm 6.3$   $\mu\text{g}/\text{liter}$ ; range: 0.07–21.0  $\mu\text{g}/\text{liter}$ ; median: 5.4  $\mu\text{g}/\text{liter}$ ) was significantly lower than that for control subjects ( $P < 0.001$ ) but did not differ significantly from the MPHD patients ( $P = 0.051$ ).

The cut-point to minimize misclassification of MPHD patients and control subjects (CART, 96% sensitivity and 92%

specificity) and the ROC analysis cut-point affording 95% sensitivity (with 92% specificity) were virtually identical at 5.1 and 5.0  $\mu\text{g}/\text{liter}$ , respectively, as shown in Figs. 1A and Table 3. To attain a higher degree of specificity (95%), a cut-point of 3.3  $\mu\text{g}/\text{liter}$  can be employed, with a sensitivity of 89%. Among patients with 0–1 PHD, 47% had peak GH values below 5.1  $\mu\text{g}/\text{liter}$ .

The time points at which peak serum GH occurred ranged from 20 to 120 min following insulin injection, without a single time point providing adequate information in any group. The majority of the control subjects (88%) experienced a peak serum GH at the 60 or 90 min time points, whereas half of the MPHD patients did not attain a peak GH until the 120 min time point.

Five of the 68 subjects undergoing ITT had blood glucose nadirs above 40 mg/dl (range: 42–51 mg/dl); three were in the MPHD group, and there was one in each of the other two groups. However, in four of these five subjects, symptoms of hypoglycemia were present. The fifth subject was asymptomatic when the bedside blood glucose monitor read 35 mg/dl, but the simultaneously drawn laboratory value later returned at 51 mg/dl. One of these five subjects was in the control group, and as noted above, he had a peak GH response less than 0.05  $\mu\text{g}/\text{liter}$  to the ITT. All other control subjects had glucose nadirs  $\leq 40$  mg/dl. The cut-point to minimize misclassification of MPHD patients and control subjects was not altered by excluding from the analysis the one control subject and three MPHD patients with glucose nadirs above 40 mg/dl. Twenty-seven subjects received iv dextrose for treatment of symptomatic hypoglycemia. The percentage of subjects with glucose nadirs below 20 mg/dl was significantly higher (58%) in the group of subjects receiving iv dextrose (25%,  $P = 0.007$ ). There was no significant difference in the mean peak GH response to ITT between subjects who received iv dextrose during the test (40%) compared with those who did not (60%) ( $P = 0.27$ ).

Subjects undergoing the ITT reported numerous side effects, of which the most common were sweating (79%), vasodilatation or flushing (47%), hunger (40%), asthenia (38%), dizziness (34%), somnolence (29%), palpitations (21%), abnormal thinking (16%), tachycardia (15%), thirst (15%), nausea (12%), and nervousness (12%). Other side effects, which occurred in 5–10% of the subjects, were headache, speech disorder, chills, tremor, postural hypotension, amblyopia, and sleep disorder. Although most of these events were rated “mild” or “moderate” in severity, the intensity of these events was considered to be “severe” in a small number of patients, including one episode of syncope. However, there were no serious adverse events requiring hospitalization or causing permanent sequelae.

#### ARG-GHRH test

The combination of ARG infusion and GHRH bolus produced a mean peak GH response in the patients with MPHD of  $1.4 \pm 1.7$   $\mu\text{g}/\text{liter}$  (range: 0.025–7.7  $\mu\text{g}/\text{liter}$ , median: 0.95  $\mu\text{g}/\text{liter}$ ). Ninety-five percent of these subjects had peak serum GH levels below 5  $\mu\text{g}/\text{liter}$  and 87% had levels below 3  $\mu\text{g}/\text{liter}$ .

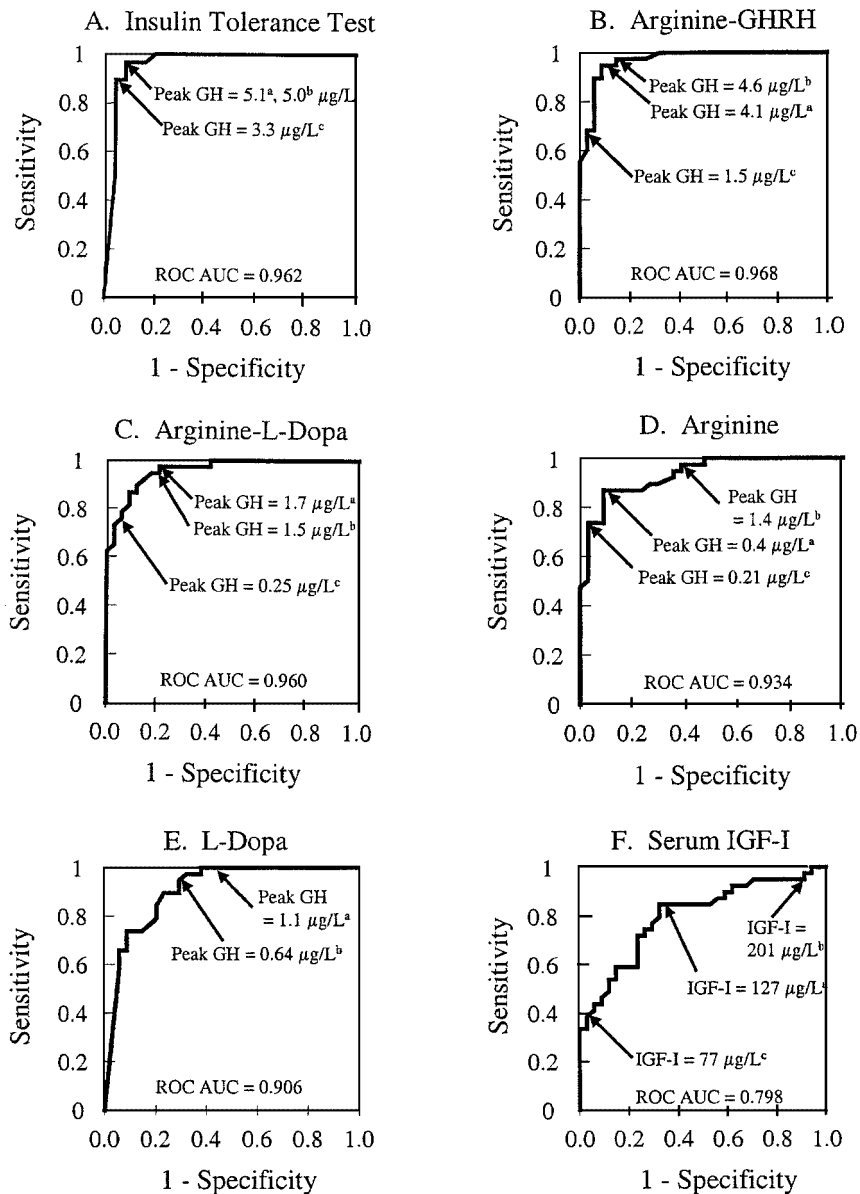


FIG. 3. Receiver-operating characteristic (ROC) curves for peak serum GH responses to: A, ITT; B, ARG-GHRH; C, ARG-L-DOPA; D, ARG; E, L-DOPA and for serum IGF-I (panel F). The ROC curve plots the true positive rate (sensitivity) against the false-positive rate (1-specificity) for different cut-points. A test with perfect discrimination between MPHD patients and matched control subjects (100% sensitivity and 100% specificity) would coincide with the *left upper corner* of the box, and be associated with a ROC area of 1.0. In contrast, a test providing no discrimination between groups would result in a diagonal line from the *left lower* to the *right upper corner* of the box (sensitivity = 1-specificity), and correspond to a ROC area of 0.5. The *arrows* and *superscripts* indicate the location on the ROC curves of the three diagnostic cut-points (see Table 3) defined as follows: a, minimize misclassification of MPHD patients and control subjects; b, 95% sensitivity for GHD; and c, 95% specificity for GHD. The area under the curve (AUC) for each ROC curve is shown.

In the control subjects, the peak GH response to ARG-GHRH was significantly ( $P < 0.001$ ) higher than in the MPHD patients (mean:  $18.4 \pm 22.1$   $\mu\text{g}/\text{liter}$ ; range: 1.2–127.0  $\mu\text{g}/\text{liter}$ ; median: 12.0  $\mu\text{g}/\text{liter}$ ). Two control subjects (6%) had peak serum GH levels less than 3  $\mu\text{g}/\text{liter}$  and five (15%) had peak levels less than 5  $\mu\text{g}/\text{liter}$ . The two male control subjects who had peak levels less than 3  $\mu\text{g}/\text{liter}$  did not differ substantially from the rest of the group in age or BMI. One of these men also had no GH response to an ITT, as reviewed above. Both subjects had a normal response to at least two other tests, as defined by CART analysis.

In the patients with a history of pituitary disease and 0–1 PHD, the peak GH response (mean:  $9.6 \pm 6.3$   $\mu\text{g}/\text{liter}$ ; range: 0.51–20.0  $\mu\text{g}/\text{liter}$ ; median: 11.0  $\mu\text{g}/\text{liter}$ ) was significantly lower than that for control subjects ( $P = 0.02$ ) but was higher than that for the MPHD patients ( $P = 0.03$ ).

The cut-point to minimize misclassification of MPHD pa-

tients and control subjects (CART, 95% sensitivity, 91% specificity) and the ROC analysis cut-point affording 95% sensitivity (with 85% specificity) were nearly identical at 4.1 and 4.6  $\mu\text{g}/\text{liter}$ , respectively, as shown in Fig. 1B and Table 3. For a specificity of 95%, a cut-point of 1.5  $\mu\text{g}/\text{liter}$  yielded a sensitivity of 68%. Among patients with 0–1 PHD, 29% had peak GH values less than 4.1  $\mu\text{g}/\text{liter}$ .

Following ARG-GHRH, the time points at which peak serum GH occurred ranged from 30 to 120 min; no single time point provided adequate information in any group. The majority of the control subjects (82%) experienced a peak serum GH at the 30 or 60 min time points, whereas the majority (68%) of the MPHD patients attained a peak GH between the 60 and 90 min time points.

The most common side effect reported during the ARG-GHRH test was vasodilatation or flushing, which was seen in 58% of subjects. Other side effects, which occurred in

**TABLE 3.** Peak GH and serum IGF-I cut-points<sup>a</sup> for the diagnosis of adult GHD

Test	Minimize misclassification (CART)				ROC AUC	95% Sensitivity (ROC)			95% Specificity (ROC)		
	Cut-point <sup>b</sup>	PPV	Sensitivity	Specificity		Cut-point <sup>b</sup>	PPV	Specificity	Cut-point <sup>b</sup>	PPV	Sensitivity
ITT	5.1	93%	96%	92%	0.962	5.0	93%	92%	3.3	96%	89%
ARG-GHRH	4.1	92%	95%	91%	0.968	4.6	88%	85%	1.5	96%	68%
ARG-L-DOPA	1.7	84%	97%	79%	0.960	1.5	85%	79%	0.25	96%	75%
ARG	0.4	92%	87%	91%	0.934	1.4	75%	62%	0.21	97%	74%
L-DOPA	1.1	75%	100%	62%	0.906	0.64	78%	71%	N/A	N/A	N/A
IGF-I	127.1	75%	85%	68%	0.798	200.7	54%	9%	77.2	94%	40%
IGF-I SDS	−2.00	100%	46%	100%	0.790	0.12	54%	6%	−1.94	95%	49%

<sup>a</sup> Three cut-points were determined for each test, derived from comparison of results obtained in patients with MPHD and matched control subjects: 1) to minimize misclassification of control subjects and MPHD patients (balance between high sensitivity and high specificity); 2) 95% sensitivity for GHD; and 3) 95% specificity for GHD. The PPV, sensitivity and specificity for the three cut-points are shown for each test, along with the ROC AUC for each test.

<sup>b</sup> Units are  $\mu\text{g/liter}$  for all cut-points except the IGF-I SDS cut-points, which have no units. AUC, Area under the curve.

5–10% of the subjects, were paresthesias, nausea, and abnormal taste sensation.

#### ARG-L-DOPA test

The combined ARG-L-DOPA test produced a mean peak GH response in the patients with MPHD of  $0.31 \pm 0.65 \mu\text{g/liter}$  (range: 0.025–3.5  $\mu\text{g/liter}$ , median: 0.05  $\mu\text{g/liter}$ ). All but one of these subjects had a peak serum GH level less than 3  $\mu\text{g/liter}$ .

The peak GH response in the control subjects was significantly ( $P < 0.001$ ) higher than in the MPHD patients (mean:  $6.7 \pm 7.1 \mu\text{g/liter}$ ; range: 0.16–37.0  $\mu\text{g/liter}$ ; median: 5.4  $\mu\text{g/liter}$ ). However, a substantial number of control subjects (35%) had a peak serum GH level less than 3  $\mu\text{g/liter}$ , and 47% of this group had values less than 5  $\mu\text{g/liter}$ .

In patients with 0–1 PHD, the peak GH response (mean:  $3.0 \pm 2.4 \mu\text{g/liter}$ ; range: 0.09–8.8  $\mu\text{g/liter}$ ; median: 3.2  $\mu\text{g/liter}$ ) was significantly lower than that for control subjects ( $P = 0.004$ ) but was higher than that for the MPHD patients ( $P = 0.04$ ).

A cut-point of 1.7  $\mu\text{g/liter}$  minimized the misclassification of MPHD patients and control subjects and provided 97% sensitivity and 79% specificity (Fig. 1C and Table 3). Using sensitivity set at 95% with ROC analysis, a cut-point of 1.5  $\mu\text{g/liter}$  yielded 79% specificity. To achieve a specificity of 95%, a low cut-point of 0.25  $\mu\text{g/liter}$  was required, resulting in a sensitivity of 75%. Among patients with 0–1 PHD, 40% had peak GH values less than 1.7  $\mu\text{g/liter}$ .

The time to peak serum GH occurred at 30 or 60 min in 68% of the control subjects. In contrast, peak serum GH was measured at the 120 min time point in 68% of the MPHD patients.

The most common side effects during the ARG-L-DOPA test were nausea (29%), vomiting (12%) and paresthesias (12%). Other side effects reported in 5–10% of subjects included asthenia, dizziness, abnormal taste sensation, and dry mouth.

#### ARG test

Following ARG infusion, the mean peak serum GH response in the patients with MPHD was  $0.3 \pm 0.51 \mu\text{g/liter}$  (range: 0.025–2.4  $\mu\text{g/liter}$ , median: 0.09  $\mu\text{g/liter}$ ). All patients had a peak serum GH level less than 3  $\mu\text{g/liter}$ .

In control subjects, the peak GH response was significantly ( $P < 0.001$ ) higher than in the MPHD patients (mean:  $3.7 \pm 3.3 \mu\text{g/liter}$ ; range: 0.08–11.0  $\mu\text{g/liter}$ ; median: 2.6  $\mu\text{g/liter}$ ). However, the majority of control subjects (59%) had peak serum GH levels less than 3  $\mu\text{g/liter}$  and 68% of this group had values less than 5  $\mu\text{g/liter}$ .

In the group of patients with 0–1 PHD, the peak serum GH response (mean:  $2.6 \pm 2.5 \mu\text{g/liter}$ ; range: 0.06–9.5  $\mu\text{g/liter}$ ; median: 2.1  $\mu\text{g/liter}$ ) was significantly higher than that for the MPHD patients ( $P < 0.001$ ) but did not differ significantly from the control subjects ( $P = 0.12$ ).

As shown in Fig. 1D and Table 3, a cut-point of 0.4  $\mu\text{g/liter}$  minimized the misclassification of MPHD patients and control subjects and produced a sensitivity of 87% and a specificity of 91%. For a higher sensitivity (95%), a cut-point of 1.4  $\mu\text{g/liter}$  resulted in a specificity of 62%. To achieve a high specificity (95%), a lower cut-point of 0.21  $\mu\text{g/liter}$  was required, which was associated with a sensitivity of 74%. Among patients with 0–1 PHD, 19% had peak GH values less than 0.4  $\mu\text{g/liter}$ .

Sixty-five percent of control subjects had a peak serum GH level measured at the 60 or 90 min time points, whereas the peak occurred at 90 or 120 min in 74% of the MPHD patients.

Side effects were uncommon with the ARG test, but 5–10% of subjects reported paresthesias, dry mouth, and headache.

#### L-DOPA test

The mean peak serum GH response to L-DOPA in the patients with MPHD was  $0.13 \pm 0.22 \mu\text{g/liter}$  (range: 0.025–1.0  $\mu\text{g/liter}$ , median: 0.03  $\mu\text{g/liter}$ ). None had peak serum GH levels above 3  $\mu\text{g/liter}$ .

In control subjects, the peak GH response was significantly ( $P < 0.001$ ) higher than in the MPHD patients (mean:  $3.3 \pm 4.9 \mu\text{g/liter}$ ; range: 0.025–20.0  $\mu\text{g/liter}$ ; median: 1.4  $\mu\text{g/liter}$ ). Nevertheless, the majority of control subjects (68%) had peak serum GH levels below 3  $\mu\text{g/liter}$  and most (79%) were less than 5  $\mu\text{g/liter}$ .

In the patients with 0–1 PHD, the peak serum GH response (mean:  $1.4 \pm 1.6 \mu\text{g/liter}$ ; range: 0.025–5.3  $\mu\text{g/liter}$ ; median: 0.74  $\mu\text{g/liter}$ ) was significantly lower than that for control subjects ( $P = 0.03$ ) but did not differ significantly from the MPHD patients ( $P = 0.13$ ).

A cut-point of 1.1  $\mu\text{g/liter}$  minimized the misclassification

of MPHD patients and control subjects and provided 100% sensitivity and a specificity of 62%, as shown in Fig. 1E and Table 3. A lower cut-point of 0.64  $\mu\text{g}/\text{liter}$  increased the specificity to 71%, while maintaining 95% sensitivity. However, ROC analysis demonstrated that it was not possible to reach 95% specificity with this test. Among patients with 0–1 PHD, 55% had peak GH values less than 1.1  $\mu\text{g}/\text{liter}$ .

The peak serum GH was measured at 90 or 120 min in 74% of control subjects and in 87% of MPHD patients.

Twenty-six percent of subjects reported nausea with L-DOPA. Other side effects seen in 5–10% of subjects were dizziness, asthenia, and headache.

#### Comparison of GH stimulation test results

Results obtained with the five GH stimulation tests were compared using the cut-points to minimize misclassification of MPHD patients and control subjects (CART analysis) for each test. Specifically, a patient was considered to have a normal response if the peak GH value exceeded the test-specific cut-point. Only two MPHD patients responded normally on more than one test. Four other MPHD patients responded normally to one test but not to the other tests. Only one MPHD patient, who had a pituitary adenoma, responded normally to the ITT but subsequently withdrew from the study and did not have any other tests. Of these seven patients, six had pituitary adenomas and 1 had a craniopharyngioma. The patient with the craniopharyngioma responded normally to ARG-GHRH and ARG but not to the ITT (peak GH = 0.43  $\mu\text{g}/\text{liter}$ ), L-DOPA, or ARG-L-DOPA. The only other MPHD patient who responded normally to ARG-GHRH had a pituitary adenoma and had abnormal responses to ARG-L-DOPA, ARG, and L-DOPA; he did not have an ITT due to his age (63.5 yr).

#### Effect of sex, age, and BMI on peak GH values

For the GH response to ITT in control subjects, BMI had a significant inverse relationship with peak serum GH when controlling for age ( $r = -0.43$ ,  $P = 0.034$ ); similar results were observed with ARG-GHRH ( $r = -0.36$ ,  $P = 0.037$ ). For every increase of 1  $\text{kg}/\text{m}^2$ , the peak serum GH level was 0.89  $\mu\text{g}/\text{liter}$  lower for the ITT and 1.4  $\mu\text{g}/\text{liter}$  lower for the ARG-GHRH test. In contrast, BMI had no significant effect on the peak GH response to the other three stimulation tests in the control subjects. Among MPHD patients, BMI had an inverse effect on peak GH when controlling for age with L-DOPA ( $r = -0.32$ ,  $P = 0.05$ ) but no significant effect with the other stimulation tests. Age had no significant effect on peak serum GH in control or MPHD subjects for any of the five GH stimulation tests.

Sex had a significant effect on peak GH in control subjects when controlling for age with the ARG ( $P = 0.01$ ), and ARG-GHRH tests ( $P = 0.03$ ). The female control subjects ( $n = 4$ ) had higher peak serum GH values than the male control subjects ( $n = 20$ ) on the ARG test ( $5.2 \pm 3.3$  vs.  $2.6 \pm 3.0$   $\mu\text{g}/\text{liter}$ ), and on the ARG-GHRH test ( $27.1 \pm 30.5$  vs.  $12.3 \pm 10.9$   $\mu\text{g}/\text{liter}$ ). There was no statistically significant effect of sex on the GH responses to the other tests in control subjects, or on the results of any stimulation test in the patients with MPHD.

Despite these relationships identified by regression analyses, the inclusion of sex, age, and BMI in the model did not alter the peak GH cut-points arrived at by CART analysis.

#### Patient test preferences

Table 4 shows subject preferences for the five stimulation tests. The ITT was the least preferred test by subjects in all three study groups. ARG alone was ranked as the most preferred test. The ARG-GHRH test was preferred significantly more than the ITT ( $P = 0.001$ ).

#### IGF-I

The serum IGF-I concentration at the screening visit was highly correlated with the level at the first stimulation test visit ( $r = 0.87$ ,  $P < 0.001$ ). Correlations were similar among the subgroups (control subjects:  $r = 0.80$ ; MPHD patients:  $r = 0.87$ ; 0–1 PHD patients:  $r = 0.89$ ,  $P < 0.001$  for all). There were no significant differences between the IGF-I concentrations at visit 1 compared with visit 2 for control subjects ( $141.6 \pm 44.6$  vs.  $133.1 \pm 35.9$   $\mu\text{g}/\text{liter}$ ,  $P = 0.074$ ), MPHD patients ( $90.9 \pm 45.6$  vs.  $92.7 \pm 48.5$   $\mu\text{g}/\text{liter}$ ,  $P = 0.66$ ) or 0–1 PHD patients ( $127.1 \pm 58.4$  vs.  $117.2 \pm 62.3$   $\mu\text{g}/\text{liter}$ ,  $P = 0.14$ ). Therefore, the screening visit IGF-I values were used in the evaluation of IGF-I as a diagnostic test using CART and ROC analyses. When all three groups were considered together, there were statistically significant ( $P < 0.05$ ) correlations between peak GH and serum IGF-I for each GH stimulation test ( $r$  values ranging from 0.21 to 0.44), but this relationship only accounted for 4–19% of the variance in the serum IGF-I values. Within each group of patients, the correlations between peak GH and IGF-I were less consistent with relatively small sample sizes.

Although the mean serum IGF-I levels were significantly different in MPHD patients compared with carefully matched control subjects ( $P < 0.001$ ), there was substantial overlap between the two groups (Fig. 1F). In the patients with a history of pituitary disease and 0–1 PHD, serum IGF-I concentrations were intermediate between the values for the other two groups. These values were significantly higher than those observed in the MPHD patients ( $P = 0.007$ ), but they did not differ from the control group values.

Among control subjects, 3 of 34 subjects (8.8%) had IGF-I values that were slightly below the normal range for age and sex (provided by the central laboratory) at the first office visit. All three of these control subjects had low normal IGF-I

**TABLE 4.** Patient preferences among the five GH stimulation tests

	Patients with MPHD	Control subjects	Patients with 0–1 PHD
ARG <sup>a</sup>	2.0 $\pm$ 1.0	1.8 $\pm$ 1.0	1.9 $\pm$ 1.2
L-DOPA <sup>a,b</sup>	2.4 $\pm$ 1.3	2.4 $\pm$ 1.2	2.5 $\pm$ 1.3
ARG-GHRH <sup>b,c</sup>	2.7 $\pm$ 1.2	2.8 $\pm$ 1.0	2.8 $\pm$ 1.1
ARG-L-DOPA <sup>c</sup>	3.2 $\pm$ 1.2	3.2 $\pm$ 1.1	3.1 $\pm$ 0.9
ITT	4.6 $\pm$ 0.8	4.8 $\pm$ 0.4	4.7 $\pm$ 0.9

A lower number indicates higher patient preference (1 = most preferred, 5 = least preferred). Data are reported for subjects who had all five tests.

<sup>a,b,c</sup> Superscripts indicate tests with rankings that were not statistically different ( $P > 0.05$ ).



values at the second visit. Fifty-six percent of the MPHD patients had IGF-I values that were below the normal range for age and sex. Among patients with 0–1 PHD, 38.1% had IGF-I values that were below the normal range.

A cut-point of 127.1  $\mu\text{g/liter}$  minimized the misclassification of MPHD patients and control subjects and provided a sensitivity of 85% and a specificity of 68% as shown in Fig. 1F and Table 3. Among patients with 0–1 PHD, 57% had serum IGF-I values below 127  $\mu\text{g/liter}$ . For a higher sensitivity (95%), a cut-point of 200.7  $\mu\text{g/liter}$  resulted in a specificity of only 9%. However, a cut-point of 77.2  $\mu\text{g/liter}$  yielded 95% specificity and a sensitivity of 40%. Thus, an IGF-I level less than 77  $\mu\text{g/liter}$  may be useful for identification of patients with a very high probability of GHD. Alternatively, an IGF-I SDS of  $-2.00$  may be used, which provided 100% specificity but only 46% sensitivity for GHD (Table 3).

### Discussion

This study is the first to compare six methods of testing for GHD in adults with hypothalamic-pituitary disorders and in control subjects matched for age, sex, estrogen use, and BMI. We considered patients with MPHD to be at the greatest risk for GHD, as has been previously demonstrated (4, 20–21, 27–29). The GH response to ITT and the ARG-GHRH test produced the sharpest separation between MPHD patients and control subjects, but the ARG-GHRH test was preferred by patients. For all tests, the definition of test-specific cut-points improved the sensitivity and specificity for the diagnosis of adult GHD. Peak serum GH occurred earlier in the majority of control subjects than in patients with pituitary disease. However, there was a wide variation in the timing of peak GH in all three study groups, indicating that it is not advisable to streamline stimulation test blood sampling to one or two time points.

The diagnosis of GHD in adults is challenging because of the lack of a single specific biologic end-point, such as growth failure, which is the cardinal clinical sign in pediatric patients. Therefore, the confirmation of GHD largely rests on laboratory testing in the context of a history of childhood GHD or adult-onset hypothalamic-pituitary disease. The peak GH response to an ITT is more specific than measurement of 24-h spontaneous GH release for the diagnosis of adult GHD (30). The use of a cut-point or threshold for normal GH response to a stimulation test is arbitrary; GH secretion may not be completely absent in GHD, but may rather reflect a continuum between normal and abnormal (28). Few studies to date have evaluated different peak GH cut-points for different stimulation tests (5, 16, 18). The cut-points of 3 or 5  $\mu\text{g/liter}$ , which have been accepted as defining GHD, have often been applied in clinical practice regardless of the pharmacologic agents used. Another important variable to consider regarding the use of an arbitrary cut-point is the type of GH assay performed (18, 31). When the same serum samples are tested in different assays, there is wide variability in the absolute values reported. As a result, the classification of individual subjects as normal or GH deficient may change (32, 33).

In the current study, a single GH assay was employed, and

two different statistical methods were used to define cut-points for the diagnosis of GHD. Using ROC analysis, 95% sensitivity and 95% specificity cut-points were calculated for each test. High sensitivity cut-points maximize detection of adult GHD, whereas high specificity cut-points minimize misclassification of normal subjects as GH-deficient. Thus, clinicians can choose whether the priority for an individual patient is to attain high sensitivity or high specificity, and use corresponding cut-points. For example, in a patient with panhypopituitarism, in whom there is a very high probability of GHD (4, 20, 21, 27–29), a clinician might prefer a test with at least 95% sensitivity (limiting the chance of a false negative result), in order not to misclassify the patient as having normal GH secretion, and withhold potentially beneficial therapy. In contrast, in an asymptomatic patient with 0–1 PHD, the risk of GHD is lower (4, 20–21, 27–29). In such a patient, the goal might be high specificity (limiting the chance of a false positive test), to avoid the unnecessary use of GH replacement. However, if such a patient had symptoms compatible with GHD, the use of the 95% sensitivity cut-point might be deemed most appropriate by some clinicians. For a balance between high sensitivity and high specificity, the cut-point derived by CART analysis to minimize misclassification of MPHD and control subjects may be used.

There are many variables affecting GH secretion and responsiveness to provocative testing (17–19). GH release declines with age and is influenced by sex, estrogen use, and body composition (17). Obesity suppresses GH release, with 24-h GH levels in obese men reduced by 75% in comparison with age-matched normal weight subjects (34). Each one unit increase in BMI has been shown to be associated with a 6% decrease in 24-h GH secretion (35). The amount of abdominal visceral fat is a stronger predictor of 24-h GH release than is total percentage body fat (36). In addition, decreased responsiveness to stimulation tests such as GHRH, ITT, ARG, L-DOPA, and ARG-GHRH has been demonstrated in subjects with obesity and/or abdominal adiposity (8, 10, 17). Thus, permanent GHD due to organic hypothalamic-pituitary disease may be difficult to distinguish from the reversible blunting of GH secretion in obesity (18). These findings underscore the importance of using a control population matched for age, sex, estrogen use, and BMI when evaluating GH stimulation tests for the diagnosis of GHD. No prior study evaluating a number of GH stimulation test agents has controlled for all of these variables. In the present study, BMI was inversely related to peak serum GH responses to the ITT and ARG-GHRH test in control subjects. This relationship was not seen in MPHD subjects (likely due to their overall very low responses), nor was it seen for any of the other tests. Although the inclusion of age, sex, and BMI did not alter the cut-points arrived at by CART analysis, the present study was not designed to evaluate whether different peak GH cut-points are necessary based on age, sex, or BMI. The inclusion of control subjects matched for these variables resulted in lower peak GH cut-points for some of the tests than are presently used in clinical practice. Therefore, the use of these new cut-points will improve the specificity of diagnostic testing for adult GHD.

The ITT has been considered the diagnostic gold standard

in establishing the presence of GHD in adults (3). An advantage of the test is that it allows evaluation of the complete hypothalamic-somatotroph axis, making it useful in patients with both hypothalamic and pituitary disease. However, there are a number of disadvantages to this test. Patients with contraindications to hypoglycemia such as seizures and ischemic heart disease are not considered candidates, and patient safety requirements make it a labor-intensive procedure (37, 38). In addition, the ITT has poor reproducibility for an individual subject. Up to a 6-fold difference in peak GH has been demonstrated on different days in healthy adults undergoing ITTs, regardless of the degree of hypoglycemia (16, 39). Although some investigators have shown total separation of patients from carefully matched control subjects using the ITT (30), others have reported overlap between the groups (13). Our study did not evaluate reproducibility of the ITT, but it did demonstrate near-complete separation between control subjects and those patients considered at high risk for GHD, based on the presence of at least two other PHD. Only one control subject had a peak GH of less than 3  $\mu\text{g}/\text{liter}$  following ITT; his failure to respond was likely due to his morbid obesity and inadequate hypoglycemia. The high diagnostic accuracy of this test is demonstrated by the area under the ROC curve of 0.962 (1.00 would indicate perfect separation between the diseased and normal groups). A peak serum GH cut-point of 5.0  $\mu\text{g}/\text{liter}$  provided 95% sensitivity and 92% specificity for the diagnosis of GHD. Although the test is clearly useful for separating MPHD patients from control subjects, it was the least preferred stimulation test by patients.

The ARG-GHRH test performed equally well, as shown by an area under the ROC curve of 0.968, indicating that it provides an ideal alternative to the ITT. Subjects preferred this test to the ITT. The GH response to ARG-GHRH is independent of age, and there is less inter and intraindividual variability than with other stimulation tests. Thus, it has been considered the best diagnostic alternative to the ITT (3, 40, 41). However, the GH response to ARG-GHRH is decreased in healthy obese subjects (42). Interestingly, two studies have reported higher peak GH levels to ARG-GHRH than to ITT, a finding not seen in our study (5, 15). In the report by Ghigo *et al.* (15), the control subjects were within 15% of ideal body weight. Our control subjects had higher BMIs (mean:  $30.3 \pm 5.8 \text{ kg}/\text{m}^2$ ; highest value:  $45.6 \text{ kg}/\text{m}^2$ ) than in these previous reports, because they were intentionally BMI-matched to the MPHD patients. If the peak GH cut-point of 9  $\mu\text{g}/\text{liter}$ , suggested by previous studies, (5) were applied to our population, all patients with MPHD would be classified as GH deficient, but many of our control subjects would fall below this threshold as well. The peak serum GH cut-point that minimized the misclassification of MPHD and control subjects in the current study was 4.1  $\mu\text{g}/\text{liter}$ ; this cut-point provided 95% sensitivity and 91% specificity. This excellent separation between patients with GHD and carefully matched control subjects, coupled with the high degree of patient acceptability, suggests that the ARG-GHRH test is the best alternative to ITT in our population.

Many investigators have evaluated ARG alone as a provocative test of GH secretion in adults (11, 12, 43, 44). This

amino acid is believed to increase GH by suppressing endogenous somatostatin (11, 45, 46). The ARG test was ranked the most preferred test by our study subjects. ARG has been shown to have high intraindividual reproducibility, but normal subjects often have peak serum GH values less than 3  $\mu\text{g}/\text{liter}$ , producing overlap with patients considered to have severe GHD (44, 47). Indeed, over half of the control subjects in the current study would be misclassified as GH-deficient if this criterion were used. A peak serum GH cut-point of 1.4  $\mu\text{g}/\text{liter}$  provided 95% sensitivity, but only 62% specificity. To achieve 91% specificity a very low cut-point of 0.4  $\mu\text{g}/\text{liter}$  was required, but this resulted in a lower sensitivity (87%). The specificity of the ARG test can be substantially improved by combining it with L-DOPA. Children have been shown to have a greater response to the combination of these agents than to either alone, but there have been no controlled data evaluating this combination in adults (48). In the current study, a peak serum GH cut-point of 1.5  $\mu\text{g}/\text{liter}$  for the ARG-L-DOPA test provided 95% sensitivity and 79% specificity. Although testing with L-DOPA alone has been suggested as an effective alternative to the ITT, it was the least useful of the five GH stimulation tests we evaluated. The L-DOPA test had the lowest ROC area among the five stimulation tests, reflecting the fact that the majority of MPHD patients overlapped with control subjects. The earlier, more positive studies included only nonobese normal subjects, as young as 13 yr old, which may account for the higher serum GH responses to L-DOPA. These older studies also employed GH RIAs, producing higher absolute values than assays that are now commonly used (6, 7). Considering the ease of administration of L-DOPA, which is the only oral GH stimulation agent, this could be used as a screening test for determining which patients do not need further testing. A peak GH response greater than 1.1  $\mu\text{g}/\text{liter}$ , the cut-point to minimize misclassification of MPHD and control subjects, could be used to identify patients who most likely do not have GHD. All 16 of the control subjects with such a response to L-DOPA had peak GH responses greater than 5  $\mu\text{g}/\text{liter}$  with ITT. Patients with peak serum GH levels below this cut-point would need another test with a higher specificity, such as ITT, ARG-GHRH, or ARG-L-DOPA to confirm the diagnosis of GHD.

Although the mean serum IGF-I concentrations were significantly different in MPHD patients compared with carefully matched control subjects, there was substantial overlap between the two groups. This test had the lowest ROC area among the six diagnostic tests evaluated. This is not surprising, as numerous studies have demonstrated that some patients with GHD have serum IGF-I concentrations within the normal range, as was the case for 44% of the MPHD patients in the present study (4, 15, 16, 21, 27, 30, 47, 49–51). Because of the decline in serum IGF-I levels with age, the diagnostic utility of measuring this hormone is particularly low in older patients (15, 44). An IGF-I level below a certain cut-point might be useful for the diagnosis of GHD, especially in childhood-onset or young adult-onset GHD patients (50, 52). The most useful finding of the present study regarding IGF-I measurement was that a cut-point of 77.2  $\mu\text{g}/\text{liter}$  provided 95% specificity for GHD; only one control subject had an IGF-I value below this cut-point. Similarly,

another study reported that an IGF-I value below 84  $\mu\text{g}/\text{liter}$  had 95% PPV and 89% specificity for adult GHD (4). These findings might allow a subgroup of patients to undergo a single blood sample instead of a stimulation test. However, because different IGF-I assays may yield different results, and serum IGF-I concentrations may be decreased by a variety of causes, caution should be used in applying low IGF-I diagnostic cut-points to the diagnosis of adult GHD in clinical practice (3). SDS offer the advantage of taking into account age and sex. We found that in this population, an IGF-I SDS of  $-2.00$  provided 100% specificity for the diagnosis of GHD in the age range studied. Our study did not include normal subjects over the age of 68 yr; it will be important for future studies to evaluate the diagnostic utility of IGF-I SDS in elderly patients with pituitary disease.

The severity of pituitary disease, marked by the number of PHD present, is associated with the severity of GHD. The mean peak GH response to provocative testing and serum IGF-I concentrations decline progressively with increasing number of PHD (4, 20, 21, 27–29, 51). Two studies reported that 87–91% of patients with two or more PHD had severe GHD (20, 21). This was reproduced in the current study, as 89% of the MPHD patients had a peak serum GH less than 3  $\mu\text{g}/\text{liter}$  on the ITT. The concordance between different tests in an individual patient may be higher in patients with MPHD (28). In the current study, the group of subjects with 0–1 PHD had IGF-I and peak GH values that were intermediate between the MPHD and control subjects. Using the cut-points that minimized misclassification of MPHD patients and control subjects, the percentage of patients with 0–1 PHD that were classified as GHD varied from 19–55% depending on the GH stimulation test. With ITT and ARG-GHRH, the percentage of patients with 0–1 PHD who had peak GH less than 3  $\mu\text{g}/\text{liter}$  was 40% and 24%, respectively. This probably reflects the biological heterogeneity of this patient group and is consistent with the findings of previous studies. Thus, for the diagnostic evaluation of such patients, the use of cut-points with high specificity may be needed to avoid treating those who are not GH deficient.

Limitations of this study must be considered. Because the ITT was performed only in subjects under 55 yr of age, data regarding the GH response to hypoglycemia are not available for older participants. Borderline attainment of hypoglycemia in five subjects may have blunted the magnitude of the peak GH response to ITT. However, the peak GH response to ITT does not correlate with the degree of hypoglycemia in healthy adults (53). Furthermore, exclusion of these five subjects from the analysis did not change the peak GH cut-point for the ITT. Because of variability in GH assays (18, 31–33), the peak GH cut-points suggested here apply only to the particular assay used. Another limitation to any study evaluating tests for the diagnosis of GHD lies in the definition of the disorder. Because there is no absolute method for distinguishing between GH sufficiency and deficiency, there must be an arbitrary determination used to evaluate the chosen diagnostic tests. Finally, pulsatile and stimulated GH levels may be higher in the preovulatory and luteal phase of the menstrual cycle than during the follicular phase (54). Although female control subjects were matched to MPHD patients according to estrogen status, subjects with

regular menses were not studied at a particular phase of the cycle, which may have produced some variability between subjects.

Many pharmacologic agents other than those included in this study have been used to stimulate GH secretion. Clonidine alone is clearly not adequate, as many normal individuals have no detectable GH response, or show the same response as to placebo (12, 47). Glucagon injection is considered a useful alternative to ITT by some investigators, but reported by others as being less effective than ARG (7, 12, 55). A number of agents have been used in combination with GHRH. Pyridostigmine plus GHRH cannot be used across the life span in the same way as ARG-GHRH, because the potentiating effect of pyridostigmine on GH release declines with age (40). The newest agents used in combination with GHRH for the diagnosis of GHD are the synthetic GH-releasing peptides (GHRP). The combination of GHRH with hexarelin, GHRP-6, or GHRP-2 is well tolerated and useful diagnostically. Subsequent studies will be needed to confirm these promising initial findings and to evaluate the proposed cut-points, which are higher than with ITT (14, 56, 57).

In conclusion, this comparison of six tests for the diagnosis of GHD in adults with hypothalamic-pituitary disease and carefully matched control subjects demonstrated that the greatest diagnostic accuracy was obtained with the ITT and the ARG-GHRH test. While there was more overlap between MPHD patients and control subjects for ARG, L-DOPA and ARG-L-DOPA, test-specific cut-points were defined to improve the sensitivity and specificity of these tests. The ARG-L-DOPA test appears to be a reasonable third choice. These data indicate that it is possible to diagnose GHD in adults without performing an ITT, provided that test-specific cut-points are employed. The ARG-GHRH test represents an excellent alternative to the ITT for the diagnosis of GHD in adults.

### Acknowledgments

We thank Marie Cook, Marilyn Miller-Collistro, Rhodora Enriquez, Yvonne Francis, Amir Hamrahian, and Karen Pulaski, who assisted the investigators with collection of data for this study. We also thank several other individuals for their contributions to this study: Ray Carroll for statistical advice; Anne Klibanski, Charmian Quigley, Armando Amador, Amy Rosen, and Chien-Feng Chen for critical review of the manuscript; Mark Muhlhauser, Brian Oertel, Maurice Lunik, and Nekia Clayton for project management and technical support; Stan Holley, Rob Kulak, Laura Macione, Otis Mitchell, and Doreen Struble for field study management. We also thank the many other individuals who contributed to the conduct of this study.

Received October 15, 2001. Accepted February 14, 2002.

Address all correspondence and requests for reprints to: Beverly M. K. Biller, M.D., Neuroendocrine Unit, Bulfinch 457B, Massachusetts General Hospital, Fruit Street, Boston, Massachusetts 02114. E-mail: bbiller@partners.org.

This study was supported by Eli Lilly & Co.. The results of this study were reported in part at the Growth Hormone Research Society Conference, Göteborg, Sweden, September 7–9, 2000, and at the 7th International Pituitary Congress, Phoenix, Arizona, June 23–25, 2001.

### References

1. de Boer H, Blok GJ, Van der Veen EA 1995 Clinical aspects of growth hormone deficiency in adults. *Endocr Rev* 16:63–86
2. American Association of Clinical Endocrinologists 1998 AACE clinical prac-

- tic guidelines for growth hormone use in adults and children. *Endocr Pract* 4:165–173
3. **Growth Hormone Research Society** 1998 Consensus guidelines for the diagnosis and treatment of adults with growth hormone deficiency: summary statement of the Growth Hormone Research Society workshop on adult growth hormone deficiency. *J Clin Endocrinol Metab* 83:379–381
  4. **Hartman ML, Crowe BJ, Biller BMK, Ho KKY, Clemmons DR, Chipman JJ, on behalf of the HypoCSS Advisory Board and The U.S. HypoCSS Study Group**, 2002 Which patients do not require a growth hormone (GH) stimulation test for the diagnosis of adult GH deficiency? *J Clin Endocrinol Metab*, 87:477–485
  5. **Aimaretti G, Corneli G, Razzore P, Bellone S, Baffoni C, Arvat E, Camanni F, Ghigo E** 1998 Comparison between insulin-induced hypoglycemia and growth hormone (GH)-releasing hormone + arginine as provocative tests for the diagnosis of GH deficiency in adults. *J Clin Endocrinol Metab* 83:1615–1618
  6. **Eddy RL, Gilliland PF, Ibarra Jr JD, Thompson JQ, MacMurry Jr JF** 1974 Human growth hormone release: comparison of provocative test procedures. *Am J Med* 56:179–185
  7. **Lin T, Tucci J** 1974 Provocative tests of growth hormone release. *Ann Intern Med* 80:464–469
  8. **Vizner B, Reiner Z, Sekso M** 1983 Effect of L-DOPA on growth hormone, glucose, insulin, and cortisol response in obese subjects. *Exp Clin Endocrinol* 81:41–48
  9. **Page M, Dieguez C, Valcavi R, Edwards C, Hall R, Scanlon MF** 1988 Growth hormone (GH) responses to arginine and L-DOPA alone and after GHRH pretreatment. *Clin Endocrinol* 28:551–558
  10. **Cordido F, Dieguez C, Casanueva FA** 1990 Effect of central cholinergic neurotransmission enhancement by pyridostigmine on the growth hormone secretion elicited by clonidine, arginine, or hypoglycemia in normal and obese subjects. *J Clin Endocrinol Metab* 70:1361–1370
  11. **Koppeschaar HPF, ten Horn CD, Thijssen JHH, Page MD, Dieguez C, Scanlon MF** 1992 Differential effects of arginine on growth hormone releasing hormone and insulin induced growth hormone secretion. *Clin Endocrinol* 36:487–490
  12. **Rahim A, Toogood AA, Shalet SM** 1996 The assessment of growth hormone status in normal young adult males using a variety of provocative agents. *Clin Endocrinol* 45:557–562
  13. **Fisker S, Jørgensen JO, Ørskov H, Christiansen JS** 1998 L-Arginine and insulin-tolerance tests in the diagnosis of adult growth hormone deficiency: influence of confounding factors. *Clin Endocrinol* 48:109–115
  14. **Mahajan T, Lightman SL** 2000 A simple test for growth hormone deficiency in adults. *J Clin Endocrinol Metab* 85:1473–1476
  15. **Ghigo E, Aimaretti G, Gianotti L, Bellone J, Arvat E, Camanni F** 1996 New approach to the diagnosis of growth hormone deficiency in adults. *Eur J Endocrinol* 134:352–356
  16. **Hoeck HC, Vestergaard P, Jakobsen PE, Falhof J, Laurberg P** 2000 Diagnosis of growth hormone (GH) deficiency in adults with hypothalamic-pituitary disorders: comparison of test results using pyridostigmine plus GH-releasing hormone (GHRH), clonidine plus GHRH, and insulin-induced hypoglycemia as GH secretagogues. *J Clin Endocrinol Metab* 85:1467–1472
  17. **Hartman ML** 2000 Physiological regulators of growth hormone secretion. In: Juul A, Jørgensen JOL, eds. *Growth hormone in adults*, 2nd ed. Cambridge: Cambridge University Press; 3–53
  18. **Shalet SM, Toogood A, Rahim A, Brennan BM** 1998 The diagnosis of growth hormone deficiency in children and adults. *Endocr Rev* 19:203–223
  19. **Fisker S, Jørgensen JO, Christiansen JS** 1998 Variability in growth hormone stimulation tests. *Growth Horm IGF Res* 8 (Suppl):A31–A35
  20. **Toogood AA, Beardwell CG, Shalet SM** 1994 The severity of growth hormone deficiency in adults with pituitary disease is related to the degree of hypopituitarism. *Clin Endocrinol* 41:511–516
  21. **Bates AS, Evans AJ, Clayton RN** 1995 Assessment of GH status in adults with GH deficiency using serum growth hormone, serum insulin-like growth factor-I and urinary growth hormone excretion. *Clin Endocrinol* 42:425–430
  22. **Copeland KC, Underwood LE, Van Wyk JJ** 1980 Induction of immunoreactive somatomedin C in human serum by growth hormone: dose-response relationships and effect on chromatographic profiles. *J Clin Endocrinol Metab* 50:690–697
  23. **Stene MA, Moxness MS, Sinha YN**, High sensitivity immunoassay for growth hormone: adult normal ranges and clinical utility for predicting acromegaly with single sample basal level testing. Program of the 80th Meeting of The Endocrine Society, New Orleans, LA, 1998 (Abstract P2-181), pp 290–291
  24. **Blum WF, Brier BH** 1994 Radioimmunoassays for IGFs and IGFFBPs. *Growth Regul* 4 (Suppl 1):11–19
  25. **Mausner JS, Bahn AK** 1974 *Epidemiology*. Philadelphia: W. B. Saunders; 237–263
  26. **Vida S** 1993 A computer program for non-parametric receiver operating characteristic analysis. *Comput Methods Programs Biomed* 40:95–101
  27. **Svensson J, Johansson G, Bengtsson BÅ** 1997 Insulin-like growth factor-I in growth hormone deficient adults: relationship to population-based normal values, body composition and insulin tolerance test. *Clin Endocrinol* 46: 579–586
  28. **Lissett CA, Thompson EG, Rahim A, Brennan BM, Shalet SM** 1999 How many tests are required to diagnose growth hormone (GH) deficiency in adults? *Clin Endocrinol* 51:551–557
  29. **Colao A, Cerbone G, Pivonello R, Aimaretti G, Loche S, Di Somma C, Faggiano A, Corneli G, Ghigo E, Lombardi G** 1999 The growth hormone (GH) response to the arginine plus GH-releasing hormone test is correlated to the severity of lipid profile abnormalities in adult patients with GH deficiency. *J Clin Endocrinol Metab* 84:1277–1282
  30. **Hoffman DM, O'Sullivan AJ, Baxter RC, Ho KK** 1994 Diagnosis of growth-hormone deficiency in adults. *Lancet* 343:1064–1068
  31. **Strasburger CJ** 1998 Laboratory assessment of GH. *Growth Horm IGF Res* 8(Suppl A):41–46
  32. **Andersson AM, Ørskov H, Ranke MB, Shalet S, Skakkebaek NE** 1995 Interpretation of growth hormone provocative tests: comparison of cut-off values in four European laboratories. *Eur J Endocrinol* 132:340–343
  33. **Granada ML, Sanmarti A, Lucas A, Salinas I, Carrascosa A, Foz M, Audi L** 1990 Assay-dependent results of immunoassayable spontaneous 24-hour growth hormone secretion in short children. *Acta Paediatr Scand* 370(Suppl): 63–70
  34. **Veldhuis JD, Iranmanesh A, Ho KK, Waters MJ, Johnson ML, Lizarralde G** 1991 Dual defects in pulsatile growth hormone secretion and clearance subserve the hyposomatotropism of obesity in man. *J Clin Endocrinol Metab* 72:51–59
  35. **Iranmanesh A, Lizarralde G, Veldhuis JD** 1991 Age and relative adiposity are specific negative determinants of the frequency and amplitude of growth hormone (GH) secretory bursts and the half-life of endogenous GH in healthy men. *J Clin Endocrinol Metab* 73:1081–1088
  36. **Clasey JL, Weltman A, Patrie J, Weltman JY, Pezzoli S, Bouchard C, Thorner MO, Hartman ML** 2001 Abdominal visceral fat and fasting insulin are important predictors of 24-hour growth hormone release independent of age, gender and other physiological factors. *J Clin Endocrinol Metab* 86:3845–3852
  37. **Fish HR, Chernow B, O'Brian JT** 1986 Endocrine and neurophysiologic responses of the pituitary to insulin-induced hypoglycemia: a review. *Metabolism* 35:763–780
  38. **Jones SL, Trainer PJ, Perry L, Wass JA, Besser GM, Grossman A** 1994 An audit of the insulin tolerance test in adult subjects in an acute investigation unit over one year. *Clin Endocrinol* 41:123–128
  39. **Vestergaard P, Hoeck HC, Jakobsen PE, Laurberg P** 1997 Reproducibility of growth hormone and cortisol responses to the insulin tolerance test and the short ACTH test in normal adults. *Horm Metab Res* 29:106–110
  40. **Valetto MR, Bellone J, Baffoni C, Savio P, Aimaretti G, Gianotti L, Arvat E, Camanni F, Ghigo E** 1996 Reproducibility of the growth hormone response to stimulation with growth hormone-releasing hormone plus arginine during lifespan. *Eur J Endocrinol* 135:568–572
  41. **Ghigo E, Aimaretti G, Corneli G, Bellone J, Arvat E, Maccario M, Camanni F** 1998 Diagnosis of GH deficiency in adults. *Growth Horm IGF Res* 8(Suppl A):55–58
  42. **Ghigo E, Procopio M, Boffano GM, Arvat E, Valente F, Maccario M, Mazza E, Camanni F** 1992 Arginine potentiates but does not restore the blunted growth hormone response to growth hormone-releasing hormone in obesity. *Metabolism* 41:560–563
  43. **Merimee T, Rabinowitz D, Fineberg SE** 1969 Arginine-initiated release of human growth hormone. *N Engl J Med* 280:1434–1438
  44. **Toogood AA, O'Neill PA, Shalet SM** 1996 Beyond the somatopause: growth hormone deficiency in adults over the age of 60 years. *J Clin Endocrinol Metab* 81:460–465
  45. **Alba-Roth J, Muller OA, Schopohl J, von Werder K** 1988 Arginine stimulates growth hormone secretion by suppressing endogenous somatostatin secretion. *J Clin Endocrinol Metab* 67:1186–1189
  46. **Ghigo E, Bellone J, Mazza E, Imperiale E, Procopio M, Valente F, Lala R, De Sanctis C, Camanni F** 1990 Arginine potentiates the GHRH- but not the pyridostigmine-induced GH secretion in normal short children. Further evidence for a somatostatin suppressing effect of arginine. *Clin Endocrinol* 32: 763–767
  47. **Baum HB, Biller BMK, Katznelson L, Oppenheim DS, Clemmons DR, Cannistraro KB, Schoenfeld DA, Best SA, Klibanski A** 1996 Assessment of growth hormone (GH) secretion in men with adult-onset GH deficiency compared with that in normal men—a clinical research center study. *J Clin Endocrinol Metab* 81:84–92
  48. **Weldon VV, Gupta SK, Klingensmith G, Clarke WL, Duck SC, Haymond MW, Pagliara AS** 1975 Evaluation of growth hormone release in children using arginine and L-DOPA in combination. *J Pediatr* 87:540–544
  49. **Thissen JP, Ketelslegers JM, Maiter D** 1996 Use of insulin-like growth factor-I (IGF-I) and IGF-binding protein-3 in the diagnosis of acromegaly and growth hormone deficiency in adults. *Growth Regul* 6:222–229
  50. **Hilding A, Hall K, Wivall-Helleryd I-L, Säaf M, Melin A-L, Thorén M** 1999 Serum levels of insulin-like growth factor I in 152 patients with growth hormone deficiency, aged 19–82 years, in relation to those in healthy subjects. *J Clin Endocrinol Metab* 84:2013–2019
  51. **Marzullo P, Di Somma C, Pratt KL, Marzullo P, Di Somma C, Pratt KL, Khosravi J, Diamandis A, Lombardi G, Colao A, Rosenfeld RG** 2001 Usefulness of different biochemical markers of the insulin-like growth factor (IGF)

- family in diagnosing growth hormone excess and deficiency in adults. *J Clin Endocrinol Metab* 86:3001–3008
52. Span JP, Pieters GF, Sweep CG, Swinkels LM, Smals AG 1999 Plasma IGF-I is a useful marker of growth hormone deficiency in adults. *J Endocrinol Invest* 22:446–450
  53. Hoeck HC, Vestergaard P, Jakobsen PE, Laurberg P 1995 Test of growth hormone secretion in adults: poor reproducibility of the insulin tolerance test. *Eur J Endocrinol* 133:305–312
  54. Faria AC, Bekenstein LW, Booth Jr RA, Vaccaro VA, Asplin CM, Veldhuis JD, Thorne MO, Evans WS 1992 Pulsatile growth hormone release in normal women during the menstrual cycle. *Clin Endocrinol* 36:591–596
  55. Podolsky S, Sivaprasad R 1972 Assessment of growth hormone reserve: comparison of intravenous arginine and subcutaneous glucagon stimulation tests. *J Clin Endocrinol Metab* 35:580–584
  56. Gasperi M, Aimaretti G, Scarcello G, Corneli G, Cosci C, Arvat E, Martino E, Ghigo E 1999 Low dose hexarelin and growth hormone (GH)-releasing hormone as a diagnostic tool for the diagnosis of GH deficiency in adults: comparison with insulin-induced hypoglycemia test. *J Clin Endocrinol Metab* 84:2633–2637
  57. Popovic V, Leal A, Micic D, Koppeschaar HP, Torres E, Paramo C, Obradovic S, Dieguez C, Casanueva FF 2000 GH-releasing hormone and GH-releasing peptide-6 for diagnostic testing in GH-deficient adults. *Lancet* 356:1137–1142

---

## American Board of Internal Medicine

### 2002 Certification Examination in Endocrinology, Diabetes, and Metabolism

Late registration period: April 2–June 1, 2002

Examination date: November 6, 2002

**Important note:** The Board now offers all of its Subspecialty Certification Examinations annually.

#### 2002 ABIM Recertification Examinations in Internal Medicine, its Subspecialties, and Added Qualifications

The ABIM Recertification Program, which has been renamed the Program for Continuous Professional Development (CPD), consists of an at-home, open-book Self-Evaluation Process (SEP) and an examination that will be administered twice each year in May and November. To register for the examination, Diplomates must be enrolled in the CPD Program and be in at least year 6 of their certification cycle.

#### CPD examination administration

May 7, 2002  
November 6, 2002

#### Deadline for submission of exam registration form

March 1, 2002  
September 1, 2002

For more information and application forms, please contact: Registration Section, American Board of Internal Medicine, 510 Walnut Street, Suite 1700, Philadelphia, Pennsylvania 19106-3699. *Telephone:* (800) 441-2246 or (215) 446-3500; *Fax:* (215) 446-3590; *E-mail:* request@abim.org; *Web Site:* www.abim.org.