Comparative Validation of the Growth Hormone-Releasing Hormone and Arginine Test for the Diagnosis of Adult Growth Hormone Deficiency Using a Growth Hormone Assay Conforming to Recent International Recommendations

Philippe Chanson, Anne Cailleux-Bounacer, Jean-Marc Kuhn, Georges Weryha, Olivier Chabre, Françoise Borson-Chazot, Séverine Dubois, Caroline Vincent-Dejean, Thierry Brue, Christine Fedou, Jean-Louis Bresson, Pierre Demolis, and Jean-Claude Souberbielle*

Context: The GHRH plus arginine (GHRH+Arg) test is a promising alternative to the insulin tolerance test (ITT) for diagnosis of adult GH deficiency (AGHD).

Objectives: The objectives of the study were to validate the GHRH+Arg test for diagnosis of AGHD, using the ITT as comparator and a GH assay calibrated according to recent international recommendations, and to study the repeatability and tolerance of both tests.

Design: This was a multicenter, randomized, open-label, phase III study.

Setting: The study was conducted at 10 French university hospitals.

Subjects: Sixty-nine subjects (38 and 15 with high and low probability of GH deficiency, respectively, and 16 healthy controls) were randomized: 35 to the GHRH+Arg-GHRH+Arg-ITT test sequence and 34 to the ITT-ITT-GHRH+Arg test sequence.

Interventions: Each subject underwent three tests of GH secretion separated by 24 h or more.

Main Outcome Measures: The primary variable used for response assessments was serum peak GH response. Test results were compared with the final AGHD diagnosis.

Results: Peak GH responses in the two tests were strongly correlated. A cutoff value of 7.89 μg/liter for GHRH+Arg corresponding to 3 μg/liter for ITT was calculated. The cutoff value leading to 95% specificity with the GHRH+Arg test was measured at about 3.67 μg/liter (sensitivity 79.0%). Intermethod agreement and repeatability were high. Both tests were well tolerated. A preference for the GHRH+Arg test was expressed by 74% of subjects.

Conclusions: The GHRH+Arg test demonstrated good accuracy and repeatability, was at least as sensitive as the ITT, and was associated with better subject acceptability. The GHRH+Arg test represents a good alternative to the ITT for the diagnosis of AGHD. **(*J Clin Endocrinol Metab* 95: 3684–3692, 2010)**

The diagnosis of adult GH deficiency (AGHD) is based on the measurement of endogenous secretion of GH in response to pharmacological stimulation. Among various provocative tests of GH secretion, the insulin tolerance test (ITT) is considered the gold standard. A GH peak measurement of 3 μg/liter or less in response to the ITT is indicative of severe AGHD, and such patients are eligible for GH replacement (1–3). However, the ITT is said to be

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in U.S.A.

doi: 10.1210/jc.2010-0295 Received February 5, 2010. Accepted April 21, 2010.
First Published Online May 19, 2010
* Author Affiliations are shown at the bottom of the next page.

Abbreviations: AE, Adverse event; AGHD, adult GH deficiency; BMI, body mass index; GHD, GH deficiency; GHRH+Arg, GHRH plus arginine; IS, international standard; ITT, insulin tolerance test; VAS, visual analog scale.

poor in terms of repeatability, is associated with side effects, and is contraindicated in certain patient groups (such as those with coronary heart disease or epilepsy) (4, 5). Also, it is of little use in insulin-resistant subjects. Among alternative tests, the combined GHRH and arginine (GHRH+Arg) test is widely considered to be the most promising (4, 6). The test appears to be reproducible and better tolerated and of similar diagnostic sensitivity/specificity to the ITT, provided that appropriate cutoff values are used (6–8).

Proposed cutoff values for the GHRH+Arg test have varied in different studies (6, 7). This may be explained, in part, by differences in the control subjects enrolled in the various studies in terms of age and body mass index (BMI) because these factors may influence peak GH response. A second problem stems from variations in GH assays, which have prevented researchers using universal reference peak of GH values. Reasons for the heterogeneity of GH assays include differences in antibody specificity; interference from GH binding proteins; and matrix effects, which may influence the GH concentration measured in samples of different nature (such as heparin or EDTA plasma and serum) (9). A further problem is variation between the GH preparations used for calibration. Thus, as a first step to improved standardization of GH assays, it has been recommended that every GH assay should be calibrated against an international standard (IS; 98/574) (10). This recalibration has induced a shift (usually a decrease) in the concentrations measured using some GH assay kits, thus potentially affecting the cutoff values defining GH deficiency (GHD). The present study is the first to use a GH assay based on recommendations that were being formulated at the time.

The purpose of the study was to validate the GHRH+Arg test for the diagnosis of AGHD using the ITT as a comparator to allow physicians to assess properly whether the test, with its advantages in terms of tolerability and simplicity, could be a better alternative to ITT. Thus, our specific aim was to determine whether, for the same level of specificity (95%), the GHRH+Arg test has a sensitivity and repeatability comparable with or greater than the ITT. We did this by modeling the intraindividual relationship between the GH responses to the two stimuli (modeling study), fixing the decision threshold of the GHRH+Arg test to obtain a specificity of 95% (validation study), and studying the inter-method agreement and repeatability of the two tests.

Participants and Methods

Participants

This was a randomized, open-label, parallel-group, phase III study conducted at 10 hospitals across France between January 2004 and November 2005. Subjects were 18–60 yr of age who had either not been treated with GH or who had stopped treatment greater than 15 d before randomization. Exclusion criteria included history of coronary problems (or electrocardiographic signs indicating an ischemic pathology), cerebrovascular insufficiency or epilepsy; intolerance to arginine, GHRH, or insulin; hyperkalemia; diabetes; BMI greater than 40 kg/m^2; severe hepatic, renal, tumoral evolutive affection, or metabolic or respiratory acidosis; immunosuppression; psychiatric disorders; Parkinson's disease or Parkinsonian syndromes treated by L-dopa; treatment with drugs directly affecting the pituitary secretion of somatotropin or provoking the release of somatostatin; antimuscarinic agents; untreated hypothyroidism or treatment with antithyroid synthesis drugs; or pregnancy or breast-feeding.

To maximize the range of peak GH values, three groups of subjects were included: healthy controls (group A) and subjects with a high (group B) or low (group C) probability of being GH deficient. Subjects considered to have a high probability of GHD were those having at least one of the following criteria: tumor of the hypothalamo-pituitary region with GH insufficiency; anterior pituitary insufficiency secondary to inflammatory, infectious, or posttraumatic pathology or to a pituitary necrosis, in which pituitary functional condition had already been documented and reevaluation of GH secretion was desired; previous irradiation of the hypothalamo-pituitary region; or known childhood-onset organic GH pituitary insufficiency and with associated pituitary hormone deficiency, excluding prolactin. Subjects considered to have a low probability of GHD had at least one of the following criteria: known idiopathic isolated GHD diagnosed during childhood and a new GH secretion test desired; nonoperated microadenoma (<1 cm of diameter); or fortuitously discovered intrasellar image (*e.g.* Rathke's pouch cyst).

The study was conducted in accordance with the Declaration of Helsinki (2000) and was approved by the Ethics Committee of Paris-Sud. Each subject provided written informed consent before participation in the study.

Study procedures and assessments

After a preinclusion visit, each subject was randomized to one of two groups (see below) and underwent three diagnostic tests (carried out during a 3 d hospitalization period or three separate

Unité Mixte de Recherche S693 (P.C.), Faculté de Médecine Paris-Sud, Université Paris-Sud 11, and Institut National de la Santé et de la Recherche Médicale Unité 693 (P.C.), Le Kremlin Bicêtre F-94276, France; Service d'Endocrinologie et des Maladies de la Reproduction et Centre de Référence des Maladies Endocriniennes Rares de la Croissance (P.C.) and Unité de Recherche Clinique (P.D.), Assistance Publique-Hôpitaux de Paris Hôpital de Bicêtre, Le Kremlin Bicêtre F-94275, France; Department of Endocrinology and Institut National de la Santé et de la Recherche Médicale CIC0204 (A.C.-B., J.-M.K.), Rouen University Hospital, Rouen F-76031, France; Service d'Endocrinologie (G.W.), Centre Hospitalier Universitaire de Nancy, Hôpital Brabois, Vandoeuvre les Nancy F-54511, France; Service d'Endocrinologie (O.C.), Centre Hospitalier Universitaire de Grenoble, Hôpital Michalon, La Tronche, F-38700 France; Service de Médecine Nucléaire (F.B.-C.), Hospices Civils de Lyon, Hôpital Neurologique, Bron 7-69500, France; Service d'Endocrinologie (S.D.), Centre Hospitalier Universitaire d'Angers, Angers F-49033, France; Service d'Endocrinologie (C.V.-D.), Centre Hospitalier Universitaire de Nantes, Hotel-Dieu, Nantes F-44093, France; Service d'Endocrinologie (T.B.), Centre Hospitalier Universitaire de Marseille, Hôpital La Timone, Marseille F-13385, France; Centre Hospitalier Universitaire de Montpellier (C.F.), Hôpital Lapeyronie, Montpellier F-34295, France; Centre d'Investigation Clinique (J.-L.B., J.-C.S.), Hôpital Necker-Enfants Malades, Assistance Publique-Hôpitaux de Paris, Paris F-75015, France; and Faculté de Médecine (J.-C.S.), Centre de Recherche Croissance et Signalisation (Institut National de la Santé et de la Recherche Médicale Unité 845), Paris F-75015, France

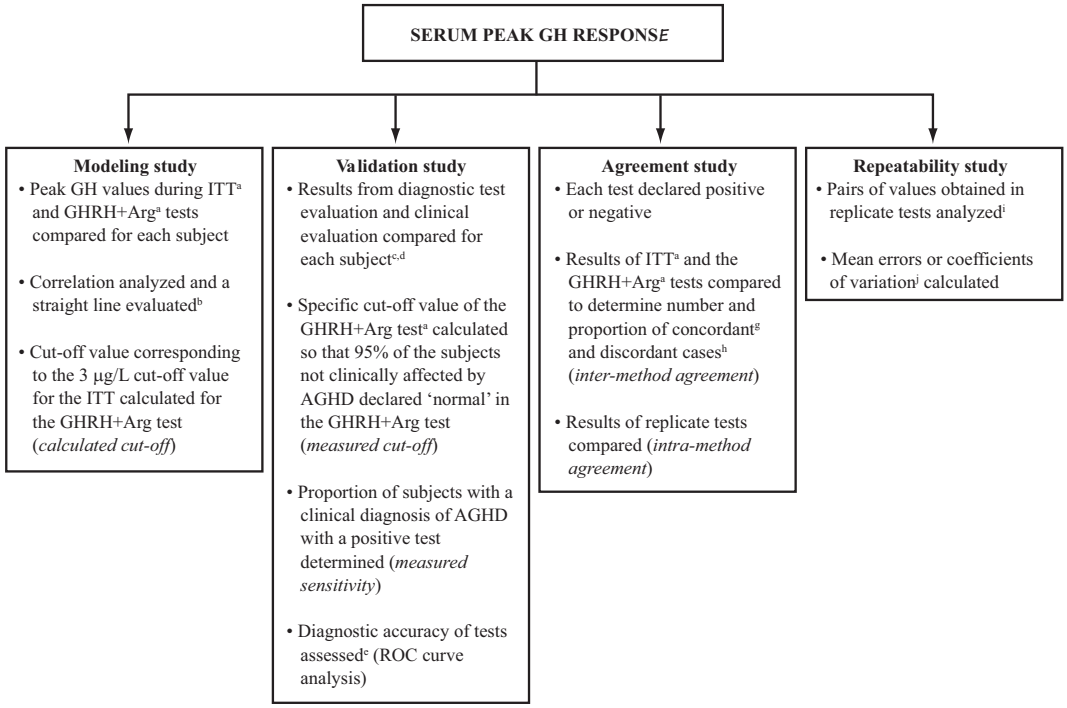


FIG. 1. Diagnostic test evaluation. a, First (or only) test. b, In the case of symmetrical distribution of the GH response in the entire modeling population (inclusive of all three groups) for at least one of the two tests. In the more likely case in which a symmetrical distribution was observed only for the healthy control group, results of modeling from this population alone could not be extrapolated to pathological subjects. c, Confirmation or rejection of the AGHD diagnosis for each subject according to standard clinical criteria, such as the presence of other pituitary deficiencies or some etiologies, *e.g.* history of craniopharyngioma, which increase the likelihood of GHD (3, 32–34). d, Uncertain cases not used. e, Method of Hanley and McNeil (35). f, ITT, cutoff value of 3 μg/liter; GHRH+Arg test, measured cutoff value. g, Same conclusion provided by two tests. h, Opposite results provided by two tests. i, Method of Bland and Altman (36). j, Depending on observed distribution. ROC, Receiver-operating characteristic.

outpatient visits), with a minimum interval of 24 h between each test. Subjects fasted for at least 10 h before each test.

The order in which the tests were administered differed according to the randomization group, improving the reliability of the data. Group 1 included the following: test 1, GHRH+Arg; test 2, GHRH+Arg; test 3, ITT. Group 2 included the following: test 1, ITT; test 2, ITT; test 3, GHRH+Arg. The test sequences were selected to assess repeatability (because each test was repeated during the first two sequences) and the agreement between tests (because both tests were completed by all subjects).

Subjects undergoing the GHRH+Arg test received GHRH (Gerel; Merck-Serono, Lyon, France), 1 μg/kg administered by an iv bolus, followed by a 30-min infusion of 0.5 g/kg arginine (Arginine Veyron; Veyron Laboratories, Marseille, France). Subjects undergoing the ITT test received an iv injection of insulin solution (Actrapid; NovoNordisk Laboratories, La Défense, France), 0.05–0.15 IU/kg. The attainment of glycemia (minimum value <40 mg/dl) was verified at each sampling time.

During both tests, blood samples were collected (at t = −15, 0, 15, 30, 45, 60, 90, and 120 min) and sent to a centralized laboratory for analysis. Serum GH levels were measured using an automated chemiluminescence-based immunometric assay (Access Ultrasensitive human GH; Beckman-Coulter, Fullerton, CA), using IS 98/574 as the GH calibrant.

At each study visit, safety assessments [adverse events (AEs), serious AEs, clinical laboratory evaluations, vital signs, and physical examinations] were performed. Subjects were also asked to give a global evaluation of tolerability for each test: using a visual analog scale (VAS), subjects were asked to rate their appreciation of the test on a scale of 0 (very acceptable) to 100 (not at all acceptable).

Diagnostic test evaluation

The primary variable used for the diagnostic test evaluation was the serum peak GH response to stimulation. The end points were the modeling of peak GH response, validation of a GHRH+Arg cutoff value, and agreement between and repeatability of the ITT and GHRH+Arg tests. The studies performed to evaluate the tests are summarized in Fig. 1.

Sample size

A sample size calculation indicated that a total of 76 subjects in groups B (high probability of GHD) and C (low probability of GHD) was necessary to guarantee, with a 5% margin of error, that specificity measured at 95% was not less than 90%. To give the study a sufficiently powerful modeling, 30 subjects in group A (healthy subjects) were also needed.

Randomization

Before the first test was performed, subjects were centrally randomized into two groups according to two test sequences using an interactive voice response system. The randomization system was designed to achieve balance between the groups with respect to age, BMI, and GHD etiology.

Statistical analysis

All randomized subjects were included in the diagnostic test evaluation. Analyses were performed on three different pre-

defined populations, depending on the end point analyzed: repeatability population (all randomized subjects who underwent the first two tests and had no major protocol deviation); modeling and agreement populations (all randomized subjects who completed all three tests and had no major protocol deviation); and validation population (exclusion of subjects with low probability of GHD).

All data were analyzed using the software SAS System version 8.2 (SAS Institute, Cary, NC).

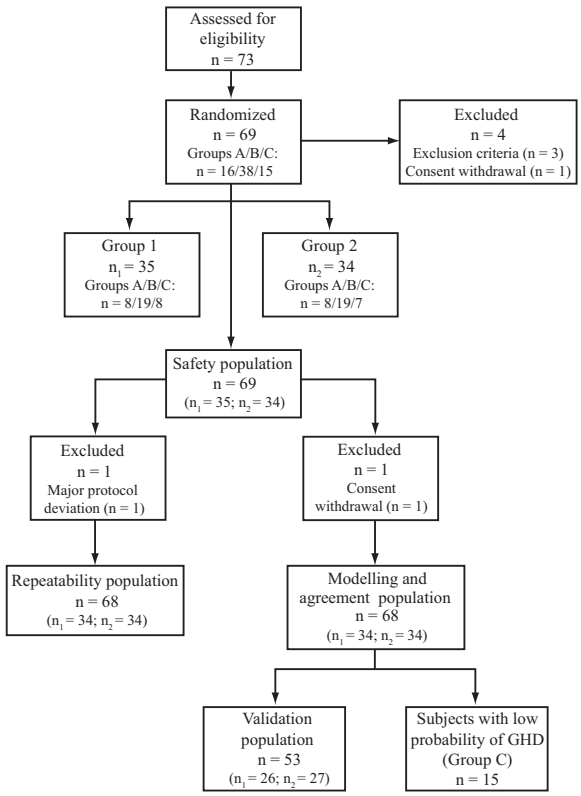
Statistical evaluations included the Student *t* and the Wilcoxon-Mann-Whitney tests (continuous variables) or the χ^2 and Fisher's exact tests (categorical and ordinal variables). All tests were two tailed and the level of significance was set at 5%.

Results

Study participants

Data were collected from 10 hospitals between January 2004 and November 2005. The study was prematurely terminated due to an interruption in the supply of GHRH.

The numbers of subjects screened and randomized to the different subgroups and included in the analysis populations are summarized in Fig. 2. A total of 69 subjects (16 in group A, 38 in group B, and 15 in group C) were randomized: 35 to group 1 and 34 to group 2. The numbers of subjects recruited to groups A and C were below the target numbers for inclusion (30 and 38, respectively).



Group 1: Test sequence = GHRH+Arg–GHRH+Arg–ITT
Group 2: Test sequence = ITT–ITT–GHRH+Arg
n_1 = number of subjects in Group 1; n_2 = number of subjects in Group 2
Group A: healthy controls; Group B: high probability of GHD; Group C: low probability of GHD

FIG. 2. Patient flow diagram.

Within each randomization group, the etiology of the GHD groups was balanced, *i.e.* groups 1 and 2 included a similar number of healthy controls and subjects with high or low probability of GHD. A total of 34 subjects (49.3%) had at least one protocol deviation (33 minor, one major).

Demographic characteristics were homogeneous within randomization and etiology of GHD groups (Table 1). Subjects were mainly male (58.0%) and white (94.2%). Some subjects presented with other pituitary hormone deficiencies indicative of hypopituitarism: 13 of 69 (18.8%) with corticotropic deficiency; 11 of 69 (15.9%) with gonadotropic deficiency; and nine of 69 (13.0%) with decreased thyrotropic function. For all patients with deficiencies, adequate replacement was stable for at least 3 months. BMI was less than 25 kg/m^2 in 25 subjects and 25 kg/m^2 or greater in 43 subjects. Apart from weight, demographic characteristics were homogeneous within BMI subgroups.

Diagnostic test evaluation

Modeling study

For the ITT, the mean first peak GH responses were 11.9, 0.9, and 8.4 μg/liter for subjects in groups A, B, and C, respectively. The corresponding values for the GHRH+Arg test were 26.1, 3.2, and 14.1 μg/liter. For both tests, the difference between the three etiological groups was significant ($P < 0.001$).

Overall, peak GH responses were higher with the GHRH+Arg test than with the ITT. With the exception of responses to the GHRH+Arg test in group A, the first peak GH responses measured in the modeling population were not normally distributed. In correlation analyses, the peak response in the GHRH+Arg test *vs.* the peak response in the ITT was plotted for each subject (Fig. 3A) and a strong positive link found ($r = 0.7$). A regression equation was used to propose the following model: GHRH+Arg peak GH value = 3.60 + (1.43 × ITT peak GH value). According to this model, the GHRH+Arg test cutoff value corresponding to the usual 3 μg/liter cutoff value for severe GHD in the ITT was 7.89 μg/liter. The relationship between the peak GH responses with the two tests in each of the three etiological groups is shown in Fig. 3B.

The first peak GH response was not significantly different when the study population was split by age and sex (data not shown). When subjects were categorized by BMI (<25 or ≥25 kg/m^2), the first GH peak in the ITT was similar between subgroups ($P = 0.211$). However, in the GHRH+Arg test, the first GH peak was significantly lower for subjects with high BMI (8.9 ng/ml for BMI ≥25 kg/m^2 *vs.* 13.7 ng/ml for BMI <25 kg/m^2; $P = 0.024$).

TABLE 1. Demographic characteristics (all randomized subjects)

	Group A (healthy controls) (n = 16)	Group B (high probability of GHD) (n = 38)	Group C (low probability of GHD) (n = 15)	Total (n = 69)
Sex, n (%)				
Male	7 (43.8)	25 (65.8)	8 (53.3)	40 (58.0)
Race, n (%)				
White	14 (87.5)	36 (94.7)	15 (100.0)	65 (94.2)
Black	0	1 (2.6)	0	1 (1.4)
Other	2 (12.5)	1 (2.6)	0	3 (4.4)
Mean (SD) age (yr)	37.8 (6.9)	42.2 (11.7)	35.4 (14.1)	39.7 (11.6)
Mean (SD) weight (kg)	70.2 (8.6)	78.1 (17.1)	68.7 (13.7)	74.2 (15.2)
Mean (SD) BMI (kg/m^2)	25.1 (1.9)	27.3 (5.2)	24.9 (5.0)	26.2 (4.7)
Etiology of GHD, n (%)				
Hypothalomo-pituitary tumor		22 (57.9)		
Secondary anterior pituitary insufficiency		7 (18.4)		
Hypothalamo-pituitary irradiation		5 (13.2)		
Organic anterior pituitary insufficiency		4 (10.5)		
Idiopathic isolated childhood-onset GHD			5 (33.3)	
Nonoperated microadenoma			3 (20.0)	
Intrasellar image			7 (46.7)	

Sex, race, age, weight, and BMI were not significantly different between GHD subgroups.

FIG. 3. Global peak GH response modeling: plot of individual peak GH response to the ITT against peak response to the GHRH+Arg test for all subjects (A): GHRH+Arg peak = 3.6043 + 1.4278 * ITT peak; R^2 = 0.484; Spearman correlation coefficient (SCC) = 0.88; and each of the three etiological groups in the modeling population (B), high probability of GHD: GHRH+Arg peak = 1.2718 + 2.527 * ITT peak; R^2 = 0.531; SCC = 0.79; low probability of GHD: GHRH+Arg peak = 6.0728 + 0.9552 * ITT peak; R^2 = 0.352; SCC = 0.43; healthy control: GHRH+Arg peak = 21.1138 + 0.4217 * ITT peak; R^2 = 0.029; SCC = 0.24.

Validation study

The cutoff value leading to a specificity of 95% (*i.e.* 95% of all subjects not clinically suffering from AGHD declared normal by the test) ranged between 3.67 and 4.96 μg/liter for the GHRH+Arg test and between 5.17 and 6.46 μg/liter for the ITT.

The sensitivity and specificity associated with different cutoff values for the two tests are shown in Table 2. Both tests recognized all healthy controls as healthy controls (specificity 100%) (if cutoff values of 3.67 μg/liter for the GHRH+Arg test and 5.17 μg/liter or 3 μg/liter for the ITT were used). However, sensitivity (the ability to identify subjects with AGHD) was higher with the ITT.

No statistical difference between the receiver-operating characteristic curves of each test was detected. The dis-

TABLE 2. Measured sensitivity and specificity with different cut points for the GHRH+Arg test and ITT (validation population)

Validation population (n = 53)	AGHD (n = 38)	Healthy controls (n = 15)	Sensitivity (%)	Specificity (%)
With a 3.67-μg/liter cutoff for GHRH+Arg test (measured cutoff)				
GHRH+Arg test positive (value <3.67 μg/liter)	30 (TP)	0 (FP)	78.95	100.00
GHRH+Arg test negative (value ≥3.67 μg/liter)	8 (FN)	15 (TN)		
With a 7.89 μg/liter cutoff for GHRH+Arg test (calculated cutoff)				
GHRH+Arg test positive (value <7.89 μg/liter)	32 (TP)	2 (FP)	84.21	86.67
GHRH+Arg test negative (value ≥7.89 μg/liter)	6 (FN)	13 (TN)		
With a 3 μg/liter cutoff for ITT (usual cutoff for severe GHD)				
ITT positive (value <3 μg/liter)	34 (TP)	0 (FP)	89.47	100.00
ITT negative (value ≥3 μg/liter)	4 (FN)	15 (TN)		
With a 5.17 μg/liter cutoff for ITT (measured cutoff for GHD)				
ITT positive (value <5.17 μg/liter)	36 (TP)	0 (FP)	94.74	100.00
ITT negative (value ≥5.17 μg/liter)	2 (FN)	15 (TN)		

FN, False negative; FP, false positive; TN, true negative; TP, true positive.

crimination ability of the ITT and the GHRH+Arg test was therefore not significantly different (data not shown).

Agreement study

In the evaluation of intermethod agreement, the GHRH+Arg test and the ITT gave identical results (*i.e.* either both gave a positive result or both gave a negative result) for 60 of the 68 subjects in the agreement population (88.2%). According to the classification of Landis and Koch (11), the agreement between the two tests was substantial [kappa value = 0.77 and significantly different from 0 ($P < 0.001$)].

In the evaluation of intramethods agreement, there was 100% agreement between the first and second GHRH+Arg tests with a cutoff value of 3.67 μg/liter.

Agreement between the first and second ITT was obtained in 31 subjects (91.2%) using a cutoff value of 3 μg/liter and 33 subjects (97.1%) using a cutoff value of 5.17 μg/liter.

According to the classification of Landis and Koch (11), the agreement between the first and the second test was deemed as almost perfect [kappa value = 1 and 0.82 for the GHRH+Arg test and the ITT, respectively, and significantly different from 0 for both tests ($P < 0.001$)].

Repeatability study

Among the subjects who had the ITT as first and second tests, the mean (SD) GH response was 5.7 (7.7) μg/liter and 6.3 (9.0) μg/liter, respectively. The corresponding figures for those who had the GHRH+Arg test as first and second tests were 11.4 (13.1) μg/liter and 11.4 (14.5) μg/liter, respectively. Within-subject coefficients of variation were 21.5% for the ITT and 17.4% for the GHRH+Arg test. A greater dispersion of values for the ITT was also observed in each subgroup (groups A, B, or C).

The first and second peak GH levels for each test are presented in Fig. 4A. A better repeatability of the GHRH+Arg test compared with the ITT was evident, particularly at low GH peak values.

Results of the repeatability analysis are presented in Fig. 4B. The overall repeatability of both tests using this method was judged to be good (difference between replicate measurements <2 SD) for 91.2% of subjects in both tests. Results remained comparable across groups A, B, and C.

Safety

Adverse events

The number of subjects who reported at least one AE [38 of 69 subjects (55.1%) during a GHRH+Arg test and 47 of 69 subjects (68.1%) during an ITT] was not significantly different between the two tests ($P = 0.081$). However, the number of AEs was lower during the GHRH+Arg tests [75 of 183 events (41.0%) *vs.* 108 of 183 events (59.0%) in the ITTs].

A greater proportion of subjects experienced AEs that were considered to be related to the test product for the GHRH+Arg test [34 of 69 subjects (49.3%) compared with 13 of 69 subjects (18.8%) for the ITT ($P < 0.001$)]. AEs considered to be test product related that occurred in more than 5% of subjects included hot flush, feeling hot, dysgeusia, and headache (GHRH+Arg test) and hyperhidrosis (ITT test).

One serious AE occurred: after insulin injection, the subject presented with a general malaise and loss of consciousness. This was considered to be probably related to the test. The subject recovered on receiving treatment with 30% glucose and completed the study.

One AE led to the dropout of a subject during the study: mild paresthesia lasted from the first to the second test (both were GHRH+Arg tests) and was not considered to be related to treatment.

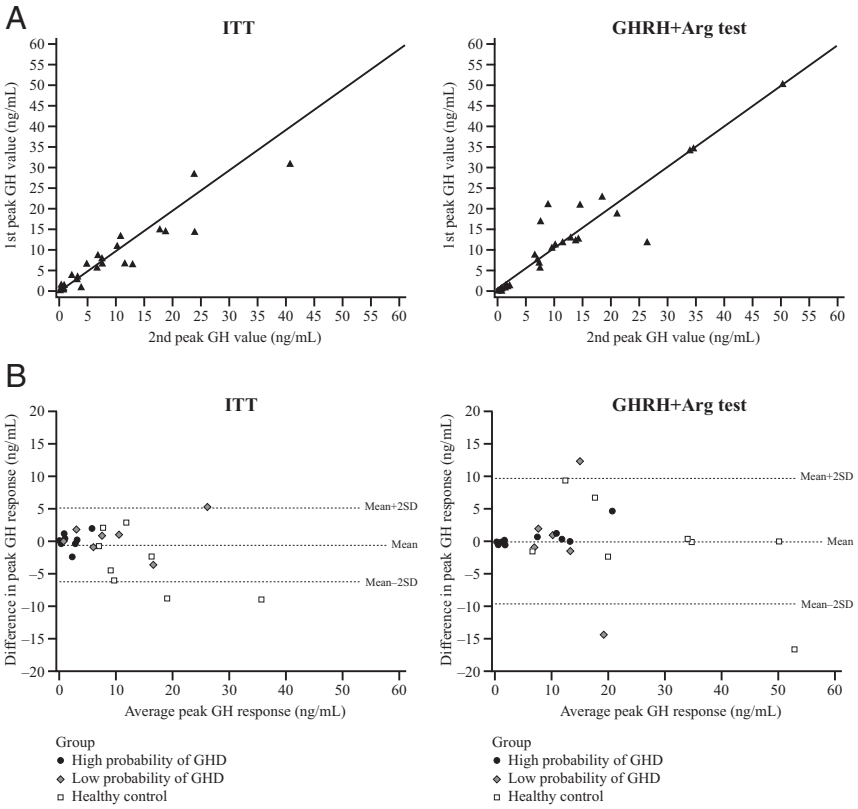


FIG. 4. Test repeatability: first and second peak GH response to the ITTs and the GHRH+Arg tests (A); Bland-Altman analysis of repeatability (repeatability population) (B).

Acceptability

Acceptance for the GHRH+Arg test was higher than for the ITT, as demonstrated by lower scores on the VAS. The mean (SD) evaluations for the GHRH+Arg tests were 14.2 (12.6) at visit 1, 16.8 (19.8) at visit 2, and 11.7 (17.6) at visit 3. For the ITTs, mean (SD) evaluations at visits 1, 2, and 3 were 25.4 (23.3), 24.5 (19.6), and 35.7 (26.9), respectively. At each visit, the VAS score was significantly higher for the ITT than the GHRH+Arg test ($P < 0.05$). Overall, 74.2% of responders (49 of 66) gave the GHRH+Arg test a lower VAS score than the ITT.

Discussion

This multicenter study was designed to validate the diagnostic stimulation test for AGHD based on the administration of GHRH and arginine, using the ITT as a comparator. Cutoff values for the diagnosis of AGHD were both calculated (modeling study) and measured (validation study). The tests were judged to have comparable accuracy, with little difference in discrimination ability, and good intermethod agreement was observed. The repeatability of the GHRH+Arg test was good, with better intramethod agreement than the ITT.

Overall, the GHRH+Arg test induced larger peak GH responses than the ITT. Accordingly, absolute values of intrapair differences appear larger for the GHRH+Arg test and the GHRH+Arg method may be considered more repeatable than the ITT; this conclusion is supported by the coefficients of variation for the two tests.

The quality of the model proposed here ($R^2 = 0.484$) was questionable because not all of the assumptions of linear regression were met: the peak GH value for the overall population did not follow a normal distribution, and the variance around the regression line was not equal across the data. The measured cutoff value of 3.67 μg/liter was therefore probably more reliable than the calculated value. The cutoff values measured in the present study are close to those obtained in previous studies (6, 7).

In the current study, the lower value for a specificity of 95% was 5.17 μg/liter for the ITT. This is in agreement with previous studies, which demonstrate that a cutoff value of 3 μg/liter defines severe GHD, whereas a cutoff value of 5 μg/liter defines GHD. Using receiver-operating characteristic curve analysis, Biller *et al.* (6) reported that balance between high sensitivity and high specificity was achieved with a cutoff level of 5.1 μg/liter. Aimaretti *et al.* (12) demonstrated that 3.8 and 5.3 μg/liter represented the first and third centile limits of the peak GH response in normal subjects.

Although the GH response to both tests was positively correlated, the GHRH+Arg test provided a more potent stimulus of GH secretion than the ITT, as reported previously (6, 7). The mean first peak GH response to the GHRH+Arg test was significantly higher for group A (26.1 μg/liter) compared with subjects in group C (14.1 μg/liter) and group B (3.2 μg/liter) ($P < 0.001$). These values were higher than those observed with the ITT (0.9 μg/liter for group B and 11.9 μg/liter for group A).

Both tests recognized all healthy controls as healthy controls (specificity 100%) if a cutoff value of 3.67 μg/liter was used for the GHRH+Arg test. However, sensitivity (ability to identify subjects with AGHD) was better with the ITT, for both a 3- and a 5.17-μg/liter cutoff value (89.5 and 94.7% *vs.* 79.0% for the GHRH+Arg test). This is because GHRH is a very potent stimulus of GH secretion, which acts on the pituitary (in contrast with hypoglycemia, which stimulates the somatotropic axis at the level of the hypothalamus). GHRH may thus be able to elicit a substantial peak GH response in patients with

partial GHD (13, 14) or radiation-induced GHD secondary to hypothalamic dysfunction, rather than pituitary dysfunction (15–17). Furthermore, it is possible that, by providing GHRH, subjects with GHD may be misclassified as normal using the GHRH+Arg test. It has previously been shown that combined administration of GHRH and GH-releasing peptide-6 was able to elicit a clear GH response in subjects for whom the ITT failed to elicit a GH response (18).

We also evaluated repeatability and agreement between tests. Agreement and repeatability of both tests were very high; this provides reassurance that the classification of a given patient as normal or abnormal is reproducible, either when the patient is tested twice using the same method or when tested using two different methods. This was somewhat unexpected for the ITT, for which intraindividual reproducibility was previously considered to be poor (19–22) but expected for GHRH+Arg because reproducibility is a well-known advantage of this test (23).

The GH response to the GHRH+Arg test was not influenced by age or sex, as demonstrated in previous studies (24, 25). However, BMI had a significant effect on the first peak GH response, with a lower value in subjects with a BMI 25 kg/m^2 or greater, as reported previously (26–30).

Both tests were well tolerated and no unexpected AEs occurred. The AEs experienced (mainly hot flush/feeling hot) have been reported in the literature (6). The high rate of AEs is likely to be related to the protocol because AEs were recorded in a manner more similar to a clinical trial than routine clinical practice. Despite a higher incidence of related AEs using the GHRH+Arg test, patients considered the tolerability of the test to be significantly better than the ITT, as has been reported previously (6). In a wider sense, the overall safety of the GHRH+Arg test is favorable, given that the ITT is contraindicated in certain patient groups.

One of the strengths of this study is that it is the first to comply with the recent European recommendations for GH assays (10). In an attempt to standardize GH assays used for diagnostic testing, a recent European consensus statement has recommended that every GH assay should be calibrated against a new international standard, IS 98/574 (10). These recommendations are to be extended worldwide in the very near future.

One potential criticism is the limited number of healthy controls, which may have decreased the power of the study. This did not prevent us performing all the descriptive statistics that were necessary; moreover, because the results are very similar to those reported already, the conclusions are unlikely to be significantly weakened. Our recruitment of healthy controls was hampered by an interruption in the supply of GHRH and the premature termination of the study. Although this form of GHRH is no longer supplied (31), similar forms are available (*i.e.* GHRH Ferring summary of product characteristic, http://emc.medicines.org.uk/document.aspx?documentId=664), and scientific institutions are trying to make the product available for routine diagnosis of AGHD.

In conclusion, results from this phase III randomized study show that the GHRH+Arg test represents a good alternative to the ITT for the diagnosis of AGHD. Furthermore, the GH cutoff limits were calculated using an assay that was calibrated according to the recent recommendations for GH assays. The GHRH+Arg diagnostic test has good accuracy and repeatability and is at least as sensitive as the ITT, provided that appropriate cutoff limits are applied. Moreover, subjects considered it more acceptable than the ITT.

Acknowledgments

We take full responsibility for the content of the paper but thank MaiLee Wong, Ph.D. (Caudex Medical supported by Merck Serono S.A., Geneva, Switzerland, an affiliate of Merck KGaA, Darmstadt, Germany) for her assistance in preparing the initial draft of the manuscript and collating the comments of the authors and other named contributors. We also thank Lory Maurin-Bernier (ICTA PM, France) for statistical input. We are also very grateful to all physicians who actively participated in the recruitment of patients, particularly Drs. Véronique Pascal-Vigneron and Georges Bruttman.

Address all correspondence and requests for reprints to: Professor Philippe Chanson, Service d'Endocrinologie et des Maladies de la Reproduction, Hôpital de Bicêtre, Assistance Publique-Hôpitaux de Paris, 78 Rue du Général Leclerc, Le Kremlin-Bicêtre 94275, France. E-mail: philippe.chanson@bct.aphp.fr.

Disclosure Summary: P.D. has nothing to declare. Institutions of all other authors received grants from Merck Serono for this study.

References

1. **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
2. **Ho KK** 2007 Consensus guidelines for the diagnosis and treatment of adults with GH deficiency II: a statement of the GH Research Society in association with the European Society for Pediatric Endocrinology, Lawson Wilkins Society, European Society of Endocrinology, Japan Endocrine Society, and Endocrine Society of Australia. Eur J Endocrinol 157:695–700
3. **Molitch ME, Clemmons DR, Malozowski S, Merriam GR, Shalet SM, Vance ML, Stephens PA** 2006 Evaluation and treatment of adult growth hormone deficiency: an Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab 91:1621–1634

4. **Gharib H, Cook DM, Saenger PH, Bengtsson BA, Feld S, Nippoldt TB, Rodbard HW, Seibel JA, Vance ML, Zimmerman D** 2003 American Association of Clinical Endocrinologists medical guidelines for clinical practice for growth hormone use in adults and children—2003 update. Endocr Pract 9:64–76
5. **Thorner MO, Bengtsson BA, Ho KY, Albertsson-Wikland K, Christiansen JS, Faglia G, Irie M, Isaksson O, Jörgensen JO, Ranke M** 1995 The diagnosis of growth hormone deficiency (GHD) in adults. J Clin Endocrinol Metab 80:3097–3098
6. **Biller BM, Samuels MH, Zagar A, Cook DM, Arafah BM, Bonert V, Stavrou S, Kleinberg DL, Chipman JJ, Hartman ML** 2002 Sensitivity and specificity of six tests for the diagnosis of adult GH deficiency. J Clin Endocrinol Metab 87:2067–2079
7. **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
8. **Ghigo E, Aimaretti G, Corneli G** 2008 Diagnosis of adult GH deficiency. Growth Horm IGF Res 18:1–16
9. **Bidlingmaier M** 2008 Problems with GH assays and strategies toward standardization. Eur J Endocrinol 159(Suppl 1):S41–S44
10. **Trainer PJ, Barth J, Sturgeon C, Wieringaon G** 2006 Consensus statement on the standardisation of GH assays. Eur J Endocrinol 155:1–2
11. **Landis JR, Koch GG** 1977 The measurement of observer agreement for categorical data. Biometrics 33:159–174
12. **Aimaretti G, Baffoni C, DiVito L, Bellone S, Grottoli S, Maccario M, Arvat E, Camanni F, Ghigo E** 2000 Comparisons among old and new provocative tests of GH secretion in 178 normal adults. Eur J Endocrinol 142:347–352
13. **Aimaretti G, Corneli G, Di Somma C, Baldelli R, Gasco V, Rovere S, Migliaretti G, Colao A, Tamburrano G, Lombardi G, Ghigo E, Camanni F** 2005 Different degrees of GH deficiency evidenced by GHRH+arginine test and IGF-I levels in adults with pituitary disease. J Endocrinol Invest 28:247–252
14. **Colao A, Di Somma C, Cuocolo A, Filippella M, Rota F, Acampa W, Savastano S, Salvatore M, Lombardi G** 2004 The severity of growth hormone deficiency correlates with the severity of cardiac impairment in 100 adult patients with hypopituitarism: an observational, case-control study. J Clin Endocrinol Metab 89:5998–6004
15. **Björk J, Link K, Erfurth EM** 2005 The utility of the growth hormone (GH) releasing hormone-arginine test for diagnosing GH deficiency in adults with childhood acute lymphoblastic leukemia treated with cranial irradiation. J Clin Endocrinol Metab 90:6048–6054
16. **Darzy KH, Aimaretti G, Wieringa G, Gattamaneni HR, Ghigo E, Shalet SM** 2003 The usefulness of the combined growth hormone (GH)-releasing hormone and arginine stimulation test in the diagnosis of radiation-induced GH deficiency is dependent on the post-irradiation time interval. J Clin Endocrinol Metab 88:95–102
17. **Ham JN, Ginsberg JP, Hendell CD, Moshang Jr T** 2005 Growth hormone releasing hormone plus arginine stimulation testing in young adults treated in childhood with cranio-spinal radiation therapy. Clin Endocrinol (Oxf) 62:628–632
18. **Leal-Cerro A, Garcia E, Astorga R, Casanueva FF, Dieguez C** 1995 Growth hormone (GH) responses to the combined administration of GH-releasing hormone plus GH-releasing peptide 6 in adults with GH deficiency. Eur J Endocrinol 132:712–715
19. **Fisker S, Jørgensen JO, Orskov H, Christiansen JS** 1998 GH stimulation tests: evaluation of GH responses to heat test versus insulin-tolerance test. Eur J Endocrinol 139:605–610
20. **Fisker S, Jorgensen JO, Orskov H, Christiansen JS** 1998 L-Arginine and insulin-tolerance tests in the diagnosis of adult growth hormone deficiency: influence of confounding factors. Clin Endocrinol (Oxf) 48:109–115
21. **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
22. **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
23. **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
24. **Ghigo E, Goffi S, Nicolosi M, Arvat E, Valente F, Mazza E, Ghigo MC, Camanni F** 1990 Growth hormone (GH) responsiveness to combined administration of arginine and GH-releasing hormone does not vary with age in man. J Clin Endocrinol Metab 71:1481–1485
25. **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
26. **Bonert VS, Elashoff JD, Barnett P, Melmed S** 2004 Body mass index determines evoked growth hormone (GH) responsiveness in normal healthy male subjects: diagnostic caveat for adult GH deficiency. J Clin Endocrinol Metab 89:3397–3401
27. **Cordido F, Alvarez-Castro P, Isidro ML, Casanueva FF, Dieguez C** 2003 Comparison between insulin tolerance test, growth hormone (GH)-releasing hormone (GHRH), GHRH plus acipimox and GHRH plus GH-releasing peptide-6 for the diagnosis of adult GH deficiency in normal subjects, obese and hypopituitary patients. Eur J Endocrinol 149:117–122
28. **Corneli G, Di Somma C, Baldelli R, Rovere S, Gasco V, Croce CG, Grottoli S, Maccario M, Colao A, Lombardi G, Ghigo E, Camanni F, Aimaretti G** 2005 The cut-off limits of the GH response to GH-releasing hormone-arginine test related to body mass index. Eur J Endocrinol 153:257–264
29. **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
30. **Qu XD, Gaw Gonzalo IT, Al Sayed MY, Cohan P, Christenson PD, Swerdloff RS, Kelly DF, Wang C** 2005 Influence of body mass index and gender on growth hormone (GH) responses to GH-releasing hormone plus arginine and insulin tolerance tests. J Clin Endocrinol Metab 90:1563–1569
31. **Yuen KC, Biller BM, Molitch ME, Cook DM** 2009 Clinical review: is lack of recombinant growth hormone (GH)-releasing hormone in the United States a setback or time to consider glucagon testing for adult GH deficiency? J Clin Endocrinol Metab 94:2702–2707
32. **Hartman ML, Crowe BJ, Biller BM, Ho KK, Clemmons DR, Chipman JJ** 2002 Which patients do not require a GH stimulation test for the diagnosis of adult GH deficiency? J Clin Endocrinol Metab 87:477–485
33. **Shalet SM, Toogood A, Rahim A, Brennan BM** 1998 The diagnosis of growth hormone deficiency in children and adults. Endocr Rev 19:203–223
34. **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 (Oxf) 41:511–516
35. **Hanley JA, McNeil BJ** 1983 A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148:839–843
36. **Bland JM, Altman DG** 1986 Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307–310