Diagnostic utility of the glucagon stimulation test in comparison to the insulin tolerance test in patients following pituitary surgery

Christian Berg, Timo Meinel, Harald Lahner, Ali Yuece, Klaus Mann and Stephan Petersenn
Department of Endocrinology, University of Duisburg-Essen, Hufelandstraße 55, 45122 Essen, Germany
(Correspondence should be addressed to C Berg who is now at Division of Endocrinology, Medical Center, University of Duisburg-Essen, Hufelandstraße 55, 45122 Essen, Germany; Email: christian.berg@uni-due.de)

Abstract

Objective: The glucagon stimulation test (GST) like the insulin tolerance test (ITT) stimulates both ACTH and GH secretion. However, there are limited data with modern assays on sensitivity and specificity for GST in comparison to ITT. The aim of this study was to evaluate the diagnostic utility of the GST for GH deficiency (GHD) and adrenal insufficiency (AI) in patients following pituitary surgery.

Design and patients: ITT and GST were performed within 7 days in 49 patients at least 3 months after transsphenoidal surgery. Serum GH and cortisol were measured by Immulite 2000 assay (Siemens AG). Receiver-operating characteristic (ROC) analysis was performed to identify the thresholds for GST.

Results: In ITT, 18/49 cases were classified as AI. ROC analysis revealed a peak cortisol value \( \geq 599 \) nmol/l in GST for adrenal sufficiency with 100% specificity and 32% sensitivity, and a peak cortisol \( < 277 \) nmol/l with \( \geq 95% \) specificity and 72% sensitivity for AI. Of the 49 subjects, 25 (51%) demonstrated levels between these cut-offs and could not be diagnosed by GST alone with sufficient accuracy. Regarding GHD, 21/49 cases were classified as insufficient by ITT. ROC analysis revealed a cut-off of 2.5 ng/ml with 95% sensitivity and 79% specificity. Of the 49 cases, seven (14%) were discordant in terms of defining GHD, with six subjects being treated for GHD according to GST although being sufficient in ITT.

Conclusion: In our prospective series of patients with pituitary disease, GST is a potential alternative test for the assessment of GH reserve, but is a poor test for ACTH reserve. Test-specific cut-offs should be applied to avoid misinterpretation.

European Journal of Endocrinology 162 477–482

Introduction

Investigation of pituitary insufficiency is an important aspect in the follow-up of patients who had undergone pituitary surgery or irradiation. Whereas patients with adrenal insufficiency (AI) may develop severe symptoms, GH deficiency (GHD) in adults typically presents with subtle signs. GHD may be clinically characterized by changes in body composition, impaired psychological well-being, reduction in bone mineral density and metabolic alterations in lipids and insulin resistance, which GH replacement reverses (1). Under these circumstances, the diagnosis in a suspected patient is purely biochemical, which is based on an insufficient GH peak response in one or two stimulation tests, depending on the number of other pituitary hormone deficiencies (2).

The insulin tolerance test (ITT) was developed in the late 1960s (3, 4), and it is the ‘gold standard’ test to determine the need for cortisol (5–7) and GH replacement (8, 9) in patients with hypothalamic-pituitary disease. It assesses the integrated central and peripheral responses to a stressful event caused by hypoglycemia. GH and cortisol responses \(< 3 \mu g/l\) and \(< 500 \) nmol/l respectively have been defined as evidence of deficiency. However, this test is unpleasant for patients, is resource-consuming, and is not without risk. It has to be performed in a specialized unit with adequate supervision, and patients with documented ischemic heart disease or seizure disorders should not undergo ITT and require alternative diagnostic tests.

The glucagon stimulation test (GST) like the ITT stimulates both the ACTH and GH secretion, and it was suggested as an alternative stimulation test in terms of efficacy. Administration of glucagon using either the s.c. or i.m. route has been shown to be effective in releasing GH and cortisol (10). However, there are no prospective data with modern assays on sensitivity and specificity for the GST in comparison to the ITT. The aim of this study was to evaluate the diagnostic utility of the GST in patients with hypothalamic–hypopituitary disease following pituitary surgery by comparing the GST to...
the ITT in order to determine the cut-off values for the GST and its sensitivity and specificity in the diagnosis of both GHD and AI.

Subjects and methods

Patients
In 2008, 49 patients (26 men, age 28–64) presented to our department at least 3 months after transsphenoidal surgery. All of them had a history of tumors in the hypothalamic or pituitary area (33 nonfunctioning adenomas, 5 prolactinomas, 5 craniopharyngiomas, and 6 meningiomas). None of these patients had received recombinant GH replacement therapy prior to the study. All patients who had pituitary hormonal deficiencies other than GHD were on appropriate replacement therapy. Patients on chronic corticosteroid replacement therapy (generally 10–15 mg hydrocortisone per day) received their last dosage at 1300 h the day before testing, resulting in a drug restriction period of at least 18 h. All patients had given informed written consent to participate in the study.

Methods
Sampling was performed during the patient’s routine clinical care for endocrinological evaluation of pituitary function after pituitary surgery. For ITT, patients received 0.15 IU/kg of regular insulin (Actrapid Novo Nordisk, Mainz, Germany) intravenously to achieve blood glucose levels below 40 mg/dl and until symptoms of hypoglycemia had developed. Blood samples for GH, cortisol, and glucose were collected at 10, 0, 15, 30, 45, 60, 90, and 120 min. In addition, all patients underwent testing with glucagon (Glucagen Diagnostic Kit, Novo Nordisk) on a separate occasion by injecting 1 mg of glucagon (1.5 mg in patients having >90 kg body weight) subcutaneously, and blood samples for GH, cortisol, and glucose were collected at −10, 0, 90, 120, 150, 180, 210, and 240 min. The maximum interval between the two dynamic tests was 7 days. Further assessment of anterior pituitary function was done by baseline hormonal testing as well as by provocative tests as required.

Serum GH levels were determined by a chemiluminescent immunometric assay (Immulite 2000 assay, Siemens AG, Erlangen, Germany). All samples from each individual patient were analyzed together. The assay was calibrated against the WHO 1st international standard (80/505) for human GH. Intra- and interassay coefficients of variation (CV) values for a low point of the standard curve were 5.4 and 7.9% respectively. For ITT, a peak GH response below 3 μg/l established the diagnosis of severe GHD according to the data given in the literature. Serum cortisol levels were determined by competitive immunoassay using commercial kits (Advia Centaur, Bayer). The analytical sensitivity of the assay was 5.5 nmol/l. Intra-assay variations as CV for various cortisol values were 3.7% (107.1 nmol/l), 3.1% (155.3 nmol/l), 2.9% (391.0 nmol/l), 3.8% (759.6 nmol/l), and 3.0% (1025.0 nmol/l) respectively. Inter-assay variations for the cortisol concentrations mentioned above were 5.5, 3.8, 3.1, 1.9, and 4.0% respectively. A peak cortisol <500 nmol/l was used to define AI. TSH deficiency was defined by low serum free thyroxine level without appropriate elevation in serum TSH. In males, secondary hypogonadism was defined by low serum testosterone with inappropriately low gonadotropin level, in premenopausal females by amenorrhea in the presence of low serum estradiol level without a rise in gonadotropin level, and in postmenopausal females by inappropriately low serum gonadotropin concentrations. Vasopressin deficiency was defined by increased serum sodium and plasma osmolality and low urinary osmolality. All other parameters were determined by routine methods.

Statistical analysis
Results (mean ±S.E.M.) are expressed as absolute values for GH and cortisol. GraphPad Prism 4.0 software for Windows (GraphPad Software, San Diego, CA, USA) was used for statistical analysis. Spearman’s rank correlation analysis was used to determine relationships between the variables. For further statistical analysis, two-way ANOVA and the Mann–Whitney and the Wilcoxon tests were performed wherever appropriate. The diagnostic accuracy of the GST in comparison to the ITT was evaluated using receiver-operating characteristic (ROC) curves to identify the thresholds for

Figure 1 Comparison of peak GH levels during ITT and GST. Individual values during ITT (circles) and GST (squares) are demonstrated. Patients were defined as GH insufficient (I) or sufficient (S) based on their GH peak response to hypoglycemia of <3 ng/ml or more than 3 ng/ml.
GST with at least 95% sensitivity for GHD and AI. In addition, the analysis was adjusted to achieve a sensitivity >95% for adrenal sufficiency (AS).

Results

Investigation of the somatotropic and the corticotropic axes by ITT

GH deficiency Insulin-induced hypoglycemia resulted in a GH peak response of more than 3 μg/l in 28/49 patients (57.1%), who were considered GH sufficient (GHS; Fig. 1). The GH mean peak (± S.E.M., range) was 11.3 μg/l (± 1.6, 3.1–32.7) in GHS patients compared to a mean peak of 0.62 μg/l (± 0.19, 0.05–2.69) in 21 GHD subjects. The GH mean peak (± S.E.M., range) was 11.3 μg/l (± 1.6, 3.1–32.7) in GHS patients compared to a mean peak of 0.62 μg/l (± 0.19, 0.05–2.69) in 21 GHD subjects. The mean age in the GHS subjects was 41.9 (± 2.4, 18–62) years and the mean body mass index (BMI) was 27.2 (± 1.2, 19.4–40) kg/m², which were not significantly different compared to a mean age of 49.1 (± 3.7, 18–79) years and a mean BMI of 27.3 (± 1.0, 18.5–36.2) kg/m² in the GHD patients. Insulin-like growth factor 1 was significantly lower in the GHD patients (91.6 ± 12.2, 25–202) than in the GHS subjects (139.2 ± 11, 62–314, P < 0.01). Among the GHS patients, mean proven dysfunction of pituitary axes (including diabetes insipidus) was 1.4 (± 0.3, 0–4) compared to 2.8 (± 0.3, 0–4, P < 0.005) in the GHD subjects.

Adrenal insufficiency In ITT, 31/49 (63.3%) subjects had a peak cortisol response > 500 nmol/l and were classified as adrenal sufficient (Fig. 2). The mean cortisol peak (± S.E.M., range) was 673.6 nmol/l (± 24.0, 515–1078) in the AS subjects compared to a mean peak of 200.1 (± 43.1, 6–496) in 18 AI subjects. The mean age in the AS subjects was 44 (± 2.4, 18–68) years and the mean BMI was 27.9 (± 1.1, 19.4–40.0) kg/m², which were not significantly different compared to a mean age of 46.7 (± 4.1, 18–79) years and a mean BMI of 25.9 (± 1.0, 18.5–31.2) kg/m² in the AI patients. Basal cortisol was significantly lower in the AI patients (185.3 ± 44, 6–387) than in the AS subjects (370.0 ± 28.11, 238–630, P < 0.01). Among the AS patients, mean proven dysfunction of other pituitary axes (including diabetes insipidus) was 1.4 (± 0.3, 0–4) compared to 3.1 (± 0.3, 0–4, P < 0.001) in the AI subjects.

Evaluation of the GST

GH deficiency The results were grouped according to the classification by ITT peak values. The GH mean peak (± S.E.M., range) was 0.62 μg/l (± 0.2, 0.1–3.7) in the GHD group versus 8.52 μg/l (± 1.6, 0.3–37.5) in the GHS group (P < 0.0001). There was a positive correlation between the GH peak responses during ITT and GST (r = 0.88, P < 0.0001) (Fig. 3a). However, GH peak values were significantly lower in the GST than in the ITT (5.1 vs 6.7 ng/ml, P < 0.01). BMI (P = 0.12) and age (P = 0.22) were not correlated to GH peak responses. Individual peaks occurred relatively later after glucagon injection compared to those induced by the ITT (137 ± 3.7, 90–180 vs 36.4 ± 1.6, 15–60 min, P < 0.0001).

ROC analysis revealed an optimal cut-off of 2.5 ng/ml with 95% sensitivity and 79% specificity for GHD (Fig. 4a). By applying this optimal threshold, 7/49
(14%) cases were discordant in terms of defining GHD when comparing the GST and ITT results. Of the 21 patients, six with normal GH response (29%) would have been treated for GHD although being sufficient in ITT.

**Adrenal insufficiency** Like for GH, cortisol results were grouped according to ITT classification. In the AI group, mean cortisol peak ($\pm$ S.E.M., range) was 213.9 nmol/l ($\pm$ 50.3, 4–582) and it was 541.5 nmol/l ($\pm$ 30.9, 252–1104) in the AS group ($P < 0.0001$). Cortisol peak responses during ITT and GST correlated positively ($r = 0.70$, $P < 0.0001$) (Fig. 3b). As for GH peak values, cortisol peak values were significantly lower in GST than in ITT (421 vs 500 nmol/l, $P < 0.001$). There was no correlation of cortisol peak responses to BMI ($P = 0.43$) or to age ($P = 0.94$). Individual peak values after glucagon injection occurred relatively later compared to the peak values induced by ITT (127 ($\pm$ 3.5, 90–180) vs 34.6 ($\pm$ 1.7, 15–60) min, $P < 0.0001$).

For the diagnosis of AI by GST, ROC analysis revealed a cut-off of 599 nmol/l with 100% sensitivity and 32% specificity (vice versa for AS: Fig. 4b). In addition, ROC analysis revealed a lower cut-off of 277 nmol/l with > 95% specificity and 72% sensitivity for AI. Of the 49 subjects, 25 (51%) demonstrated cortisol levels between these cut-offs and could not be diagnosed by GST alone with sufficient accuracy, requiring further testing. In contrast, cortisol levels above the upper cut-off and levels below the lower cut-off determined AS and AI respectively with a high probability in the remaining subjects.

Further analysis revealed that the GST could be limited to a 3-h sampling period without any loss in accuracy, as in all the subjects peak responses occurred before 180 min after glucagon injection.

**Side effects** The GST was tolerated well. The only side effects reported consisted of mild flushing and vomiting in three patients (6%) and headache in one patient (2%).

**Discussion**

In the present prospective study, we demonstrated that the GST is a potential alternative test for the assessment of GH reserve in patients with pituitary disease, but a moderately precise test for ACTH reserve. We determined new cut-off values for the currently used assays and their sensitivity and specificity in the diagnosis of adrenal and GH deficiencies.

Accurate assessment of the hypothalamic–pituitary axis is essential in patients with pituitary disease following surgery. Alternative tests other than the ITT have been sought for evaluating the hypothalamo–pituitary (HP) axis, such as arginine (11), GHRH+arginine (12–14), GH-releasing peptide-6 and combined GH-releasing peptide-6 plus GHRH (15, 16), and CRH tests (17). However, none of these tests is able to assess both cortisol and GH release in parallel. Previous studies reported that the GST is a potent and reliable stimulus of GH (18–20) and cortisol (10) secretion in healthy adults and patients. The glucagon-induced cortisol release has been shown to be ACTH dependent (21, 22). The detailed physiological mechanisms by which glucagon induces both GH and cortisol release are unclear. Some of the hypothesized mechanisms include the glycemic fluctuations during the test where blood glucose levels increase initially before decreasing later in the test (23), the generation of a peptidyl fragment associated with the GH- and ACTH-releasing activity (20), and the induction of norepinephrine secretion in stimulating GH and ACTH release via $\alpha$-receptors (10).

In a recent review by Yuen et al. the GST is recommended as the best alternative test when the ITT is not desirable or when the GHRH–ARG test is unavailable, based on the GST’s reliability and availability, its accurate and reliable discrimination between normal and true GH deficiencies, safety, and lack of influence by BMI and gender (24). Of note, the GST has only few contraindications such as underlying pheochromocytoma and insulinoma (25, 26). For GH, we found a strong positive correlation between peak responses during ITT and GST, which was comparable
to the results of Spathis et al. (21). Soliman et al. concluded previously that glucagon was at least as good as the ITT in stimulating GH release (27). Moreover, Aimaretti et al., found that although the ITT was a more effective stimulus than l-arginine, l-dopa, and clonidine, glucagon was the strongest stimulus after ITT (28). By ROC curve analysis, we found that a GH cut-off of 2.5 ng/ml is optimal to prove GH sufficiency when applying a current GH assay. Owing to this, the GST was useful for the interpretation of GH status in 71% of GH-sufficient patients, and 29% (6/21) of these patients were misdiagnosed. Hence, the GST tends to overestimate the prevalence of GHD. In comparison, Conceicao et al. determined a cut-off of 3 ng/ml for the GST when categorizing the patients by ITT using a slightly higher threshold of 5 ng/ml (29). Similar to our study, sensitivity for GHD was high with 97% showing somewhat lower specificity of 88%. Moreover, Gomez et al. suggested the same cut-off of 3 ng/ml for the GST with high sensitivity and specificity of each 100% (30). It has to be stressed that their control group did not include patients with pituitary disease, but healthy subjects. Furthermore, three different GH assays were used in that study including two different polyclonal RIA in 37% of the patients, which may give higher GH concentrations compared to the currently used immunoassays. Both test- and assay-specific cut-offs are of special importance to avoid misinterpretation, so caution is required before adopting or extrapolating cut-off values from other laboratories (31).

For cortisol, the correlation between ITT and GST peak values was somehow lower than that for GH. Cortisol peaks were significantly lower in GST than in ITT. Thereby, ROC analysis with emphasis on the high sensitivity for AI revealed a cut-off of 599 nmol/l with rather low specificity. In addition to this upper threshold, which conversely has high specificity for AS, we determined a lower threshold with high specificity for AI. Thereby, half of the patients were correctly diagnosed, leaving the other half with peak levels in the gray zone for subsequent investigation by alternative tests such as the ITT. So far, there are limited data in the literature regarding the validity of the GST in assessing hypothalamic–pituitary–adrenal function in almost half of the patients, reserving patients in the gray zone for additional testing by ITT. A correct diagnosis of GHD is obtained in almost 90% of the patients.

We conclude that the GST is a simple and safe method for stimulating GH and ACTH secretion in patients with hypothalamic–pituitary disease following pituitary surgery. It is a potential alternative test for the assessment of GH reserve, but is a poor test for ACTH reserve, as demonstrated by comparison with the ITT. Test-specific cut-offs should be applied to avoid misinterpretation.

Declaration of interest
There is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding
This research did not receive any specific grant from any funding agency in the public, commercial, or not-for-profit sector.

Acknowledgements
We acknowledge the excellent support of the personnel at the Metabolic Laboratory, Division of Endocrinology, University Hospital of Essen, Germany.

References
5 Fish HR, Chernow B & O’Brian JT. Endocrine and neurophysiologic responses of the pituitary to insulin-induced hypoglycaemia: a review. Metabolism 1980 35 763–780.


15 Petersen S, Jung R & Beil FU. Diagnosis of growth hormone deficiency in adults by testing with GHBP-6 alone or in combination with GHRH: comparison with the insulin tolerance test. European Journal of Endocrinology 2002 146 667–672.


24 Yuen KC, Biller BM, Molitch ME & Cook DM. 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? Journal of Clinical Endocrinology and Metabolism 2009 94 2702–2707.


