Growth Hormone Stimulation Test Results as Predictors of Recombinant Human Growth Hormone Treatment Outcomes: Preliminary Analysis of the National Cooperative Growth Study Database

George M. Bright, MD*; Joanne R. Julius, MS‡; John Lima, PharmD*; and Sandra L. Blethen, MD, PhD‡

ABSTRACT. Growth hormone (GH) stimulation tests are considered a prerequisite to clinical trials of recombinant human GH (rhGH) therapy, but the test results may not be predictive of the treatment outcomes with rhGH. We examined the GH stimulation test results as a predictor of the treatment outcome in a cohort of prepubertal subjects in the National Cooperative Growth Study. A standard is proposed in which a diagnosis of GH deficiency is considered appropriate when a patient has significant first-year catch-up growth and that a positive stimulation test result predicts this outcome. With this construct, a traditional interpretation of GH stimulation test results correctly identifies 64% of the rhGH treatment outcomes. The analysis shows an upper limit of diagnostic sensitivity of 82% and a lower limit of specificity of 25% in our study population. The results of our recent studies suggest that the sensitivity and specificity of the current GH stimulation tests are attributable in part to broad intersubject variation in GH clearance, rates of GH elimination, and GH volume of distribution. The combined studies suggest that the use of subject-specific pharmacokinetic parameters will improve the diagnostic interpretation of GH stimulation test results and improve rhGH treatment outcomes. Pediatrics 1999; 104:1028–1031; growth hormone stimulation tests, recombinant human growth hormone, pharmacokinetic parameters, maximal stimulated growth hormone concentration.

ABBREVIATIONS. GH, growth hormone; hGH, human GH; rhGH, recombinant human GH; NGCS, National Cooperative Growth Study; ΔHt SDS, change in height standard deviation score; GHmax, maximal stimulated GH concentration.

In the early years of growth hormone (GH) therapy, human GH (hGH) was extracted from the pituitary glands of organ donors and distributed through the National Pituitary Association.1 The demand for GH frequently exceeded the supply, and criteria were developed for determining prospectively which GH-deficient patients would most likely respond favorably to hGH treatment. Early efforts were aimed at determining whether the patient had deficient production of GH, but we have since realized that states of functional GH deficiency may arise either from a defect in GH production and secretion or from insensitivity of the GH response mechanisms.2 The pulsatile nature of GH secretion mitigated against casual blood sampling to determine the adequacy of GH production and gave rise to numerous varieties of stimulation tests of GH release. Such stimulation tests have become standard not only in endocrine practice but also for the approval of GH treatment by most health care insurers. It is not surprising that renewed attention is being given to the diagnostic accuracy of these tests.

The literature is replete with comparisons of one stimulation test with another, and no single test has achieved the status of a “gold standard.” The limitations of GH stimulation tests have been reviewed,3,4 with a consensus being that the “available methods for measuring GH secretion are neither convenient nor reliable.”5 The results of GH stimulation tests in models that have attempted to explain the outcomes of hGH treatment have not always been reliable predictive variables.6–7 Are GH stimulation tests useful or not for determining which patients will benefit
from treatment? Here, we begin to explore a novel solution to this problem.

The concentration of GH in the blood may not reflect accurately the amount of GH secreted and so not readily distinguish between states of GH sufficiency and deficiency. In earlier studies, we found that the amount of cortisol given in dexamethasone-treated subjects could not be determined accurately from changes in their plasma cortisol concentration.6,7 More recently, we have repeated these studies with injections of recombinant human GH (rhGH) in adults who have been treated with somatostatin. From these studies, we see that sufficient variation exists in the subjects’ GH volumes of distribution, elimination rates, and clearances to make determining the amount of GH that was given unlikely from measuring blood GH concentrations alone. Accordingly, more accurate determinations of the amount of GH given are possible with the use of subject-specific pharmacokinetic parameters.10 We propose that using pharmacokinetic parameters specific to each patient may eventually lead to more accurate determinations of the amount of GH secreted during a GH stimulation test and that this value will be a more reliable predictor of the response to treatment with GH.

METHODS

As a preliminary step, we examined the National Cooperative Growth Study (NCGS) database to determine the predictive ability of GH stimulation tests. We selected 236 children from substudy VI who remained prepubertal during their first year of rhGH therapy and for whom a pretreatment bone age and a midparental height were available.11 All blood samples for the GH stimulation tests were analyzed at a central laboratory. The children were selected without regard to the type of GH testing used, because many types of tests are used and we sought to determine the impact of GH stimulation testing as it is currently used. Only subjects with idiopathic GH deficiency or idiopathic short stature were included. The change in the height standard deviation score (ΔHt SDS) after 1 year of rhGH treatment was the primary indicator of the treatment outcome. We used linear regression analysis to determine the relation between the maximal stimulated GH concentration (GHmax) and the ΔHt SDS. The ΔHt SDS predicted for each subject was then calculated from the regression equation and each subject’s GHmax. The prediction residuals (actual – predicted ΔHt SDS) were then examined for dependencies on GHmax, bone age at the time of treatment, the difference between chronologic and bone ages, and midparental height. We then segregated the GH stimulation test and ΔHt SDS data in a 2 × 2 analysis scheme to determine the specificity and sensitivity of the GH stimulation test as a predictor of treatment outcome. In this scheme, the diagnosis of deficient production of GH was considered appropriate when significant first-year catch-up growth was observed. A positive stimulation test result was defined as one that predicts this outcome. Thus, a positive, or predictive, GH stimulation test result is usually associated with a low GHmax and traditionally has been referred to as a “failed” GH test. An arbitrary, but commonly used, cutoff point for GHmax of 10 μg/L was used to divide the stimulation test results as failed (<10 μg/L) or passed (≥10 μg/L). A treatment outcome of 0.5 SD in ΔHt SDS was used to divide the treatment outcomes into positive (≥0.5 SD) and negative (<0.5 SD). For example, in boys younger than 12 years, an increase of 0.5 SD in height SDS is equivalent to an increase in linear growth velocity of 1.3 to 3.7 cm per year more than that necessary to maintain their current SDS.12 We then tested the dual hypotheses that a positive GH stimulation test result (<10 μg/L) is a predictor of a positive treatment outcome and that a negative GH stimulation test result (≥10 μg/L) is a predictor of a negative outcome. The sensitivity of the procedure was then defined as the number of subjects with deficient GH production (ΔHt SDS > 0.5) and a positive GH stimulation test result (GHmax < 10 μg/L), divided by the number of subjects with deficient GH production. Specificity was defined as the number of subjects without deficient GH production (ΔHt SDS < 0.5) and a negative GH stimulation test result (GHmax ≥10 μg/L) divided by the number of subjects without deficient GH production.13

RESULTS

The patient group selected for analysis consisted of 179 boys (76%) and 57 girls (24%); 172 (73%) were considered to have idiopathic GH deficiency and 64 (27%) to have idiopathic short stature. At study entry, the mean chronologic age was 8.5 × 3.2 years, and the mean bone age was 6.2 × 3.0 years.

The regression analysis of the primary treatment indicator showed a dependency of treatment outcome on the GH stimulation test result (ΔHt SDS = 1.1–0.45 [logGHmax]; P < .001, R2 = 0.11). The percentage amount of the unexplained variance, computed as 100(1 – R2) = 89%, however, suggested that GHmax by itself would not be a strong predictor of ΔHt SDS (Fig 1). The residual differences between the actual and the predicted outcomes were of both positive and negative magnitudes, suggesting that GHmax by itself would both overpredict and underpredict the treatment outcomes. By Pearson correlation analysis, the residuals were found to decrease with greater bone age (P < .001), but were not related to height at study entry, midparental height, GHmax or the difference in chronologic and bone ages.

The segregation of patient data by treatment outcomes and stimulation test results is shown in Fig 2. This categorization makes it possible to test the hypothesis that the GH stimulation test result will predict correctly the treatment outcome. The arbitrary cutoff points of 10 μg/L for the stimulation test result and 0.5 SD in height divide the overall results into four types. In the upper left quadrant are the values for the 131 subjects (56%) in whom a failed stimulation test result was a positive predictor of significant catch-up growth; ie, the “failed” test (traditional term) was a positive predictor of the treatment outcome. In the lower right quadrant are the values for the 19 subjects (8%) in whom a passed stimulation test result predicted correctly a lack of significant catch-up growth. The GH stimulation test, as traditionally interpreted, predicted 64% of the out-

![Fig 1. Dependence of ΔHt SDS on GHmax (μg/L)](image)
comes correctly. In the lower left quadrant are the values for the 57 subjects (24%) with a failed stimulation test result who did not have significant growth, and in the upper right quadrant are the values for the 29 subjects (12%) with a passed stimulation test result who had significant growth. Thus, the stimulation test failed to predict correctly 36% of the treatment outcomes. A failed stimulation test result was associated with significant growth (ΔHt SDS > 0.5) in 131 (70%) of 188 patients, and a passed stimulation test result was associated with lack of significant growth (ΔHt SDS < 0.5) in 19 (40%) of 48 patients. Considering that GH deficiency is identified by a positive growth response and that a positive test result is a GHmax of <10 μg/L, test sensitivity can be calculated as 82% (131 / [131 + 29]) and specificity as 25% (19 / [19 + 57]). The sensitivity is an upper-limit estimate to the extent that some patients with a GHmax of ≥10 μg/L were not selected for treatment but may have grown well. The specificity is underestimated to the extent that persons who are growing well do not attend the endocrine clinic and, hence, are neither tested nor treated.

**DISCUSSION**

We have compared the results of standard GH stimulation tests with first-year rhGH treatment outcomes to test the dual hypotheses that a failed GH stimulation test is a positive predictor of good growth with rhGH treatment and a passed test will correctly predict no response. The analysis cannot be construed as entirely robust, because a number of biases can be identified, but, inasmuch as the NCGS database reflects the current standards of practice, we can determine some of the strengths and weaknesses of the GH stimulation tests as they are now being used. In our study, the outcomes in the first year of rhGH treatment were predicted correctly by a traditional interpretation of a GH stimulation test result in only two thirds of the patients.

Other caveats about interpreting GH stimulation test results have been reviewed recently. The reviews question how well the first year's treatment response predicts a beneficial effect on adult height. It is not at all certain that one type of stimulation test gives equivalent results as another, or whether the same stimulation test would give the same results when repeated. Debate continues over the relative merits of pharmacologic (stimulation) versus physiologic (sequential sampling) GH testing. The use of cutoff points for test interpretation is admittedly arbitrary.

It is possible that an additional aspect of stimulation tests diminishes their diagnostic sensitivity and specificity. We have found a poor correlation between peak plasma cortisol concentrations and the amount of cortisol given in dexamethasone-treated adults. Therefore, we have asked how accurately serum GH concentrations reflect the actual mass of GH appearing in the blood. In somatostatin-treated adults, there is sufficient intersubject variation in the GH volumes of distribution, elimination rates, and clearances to make it unlikely to determine accurately from the GH levels alone how much GH was given. Consider the error encountered in that study in predicting the amount of GH that had been given from the peak GH concentration. The difference in the actual and the predicted GH doses shows a nonrandom distribution. Therefore, the tacit assumption that GH stimulation test results are linearly and positively related to how much GH actually was secreted leads to systematic error. When the amount of GH given is small, the predicted concentration exceeds the actual concentration, producing a negative error. Thus, in persons with partial GH deficiency, there is a tendency for the GH stimulation test to give a false-normal result. Accordingly, some persons with partial GH deficiency may be overlooked for treatment. When larger amounts of GH are given, the error becomes positive; that is, the actual concentration exceeds the predicted, and some patients with adequate GH secretion therefore may be considered to have GH deficiency. This error analysis from our recent study, therefore, is in substantial agreement with our current analysis of the NCGS database. If the secretion of GH is the appropriate defining parameter for diagnosing GH deficiency (as opposed to deficient GH responsive-
ness), then trying to determine the GH concentration from GH stimulation test results will produce both positive and negative errors. If GH secretion is a valid predictor of the response to treatment with rhGH, then GH stimulation tests, as currently interpreted, will have low sensitivity and specificity.

In our study in healthy adults, we found that using subject-specific pharmacokinetic parameters results in far more accurate determinations of the amount of GH given than does using GH concentration data alone.10 We are now conducting a study in children to see whether their GH pharmacokinetic parameters will make it possible to determine GH secretion more accurately and whether the mass of secreted GH, so determined, will be a more valid predictor of rhGH treatment outcomes.

ACKNOWLEDGMENTS
This work was supported by a grant from the Genentech Foundation for Growth and Development, and by an educational grant from Genentech, Inc, South San Francisco, CA.

REFERENCES

Analysis of Bone Age Data From National Cooperative Growth Study Substudy VII

Stephen F. Kemp, MD, PhD*, and Judy P. Sy, PhD‡

ABSTRACT. National Cooperative Growth Study sub-study VII was conducted 1) to compare standardized hand–wrist and knee bone age determinations in pubertal children treated with growth hormone (GH); 2) to compare local determinations of bone ages with centrally determined bone ages; 3) to relate the response to GH therapy to the bone age determinations; and 4) to ascertain the predictive value of each type of bone age determination. Eligible subjects were those in the National Cooperative Growth Study who were at Tanner pubertal stage 2 or greater for breasts (girls) or genitals (boys). Radiographs of the hand–wrist were taken annually, and radiographs of the knee were taken at the beginning and the end of the study. Separate bone age determinations were made from these radiographs. A combined hand–wrist and knee bone age determination also was derived. There were 990 patients in the study; in 925 (677 boys), there were both hand–wrist and knee bone age determinations from the baseline pubertal radiographs. There was only one radiographic assessment in 496 patients, two in 205 patients, and three to eight in the remaining patients. The strongest correlation was between the hand–wrist bone age and the hand–wrist plus knee bone age (r = .995). Also strongly correlated were knee with hand–wrist (r = .872) and knee with hand–wrist plus knee (r = .914). For none of these bone age methods was any statistically significant difference found between the methods. The locally determined bone ages correlated strongly with the centrally determined bone ages for knee (r = .850), hand–wrist (r = .928), and hand–wrist plus knee (r = .930); however, the locally determined knee and hand–wrist values were less (by ~0.3 year) than the centrally determined values. These differences, however, do not appear to be clinically significant. Pediatrics 1999;104:1031–1036; bone age determinations, recombinant human growth hormone, hand–wrist radiographs, knee radiographs.

ABBREVIATIONS. GH, growth hormone; NCGS, National Cooperative Growth Study; SDS, standard deviation score(s).
Growth Hormone Stimulation Test Results as Predictors of Recombinant Human Growth Hormone Treatment Outcomes: Preliminary Analysis of the National Cooperative Growth Study Database
George M. Bright, Joanne R. Julius, John Lima and Sandra L. Blethen

Pediatrics 1999;104;1028

Updated Information & Services
including high resolution figures, can be found at:
/content/104/Supplement_5/1028.full.html

Subspecialty Collections
This article, along with others on similar topics, appears in the following collection(s):
Growth/Development Milestones
/cgi/collection/growth:development_milestones_sub
Endocrinology
/cgi/collection/endocrinology_sub

Permissions & Licensing
Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
/site/misc/Permissions.xhtml

Reprints
Information about ordering reprints can be found online:
/site/misc/reprints.xhtml
Growth Hormone Stimulation Test Results as Predictors of Recombinant Human Growth Hormone Treatment Outcomes: Preliminary Analysis of the National Cooperative Growth Study Database

George M. Bright, Joanne R. Julius, John Lima and Sandra L. Blethen

*Pediatrics* 1999;104;1028

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/104/Supplement_5/1028.full.html