ORIGINAL ARTICLE

Growth hormone (GH) provocation tests and the response to GH treatment in GH deficiency

T J Cole, P C Hindmarsh, D B Dunger

Arch Dis Child 2004;**89**:1024–1027. doi: 10.1136/adc.2003.043406

See end of article for authors' affiliations

Correspondence to: Professor T J Cole, Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK; tim.cole@ich.ucl.ac.uk

Revised version received 5 March 2004
Accepted for publication 18 March 2004

Objective: To identify factors, particularly the growth hormone (GH) provocation test result, affecting growth response to GH treatment in children with GH deficiency (GHD).
Subjects: A total of 337 prepubertal GHD patients aged <10 years from the UK Pharmacia KIGS database (GH response to provocation test <20 mU/l).
Outcome measure: Annual change in height standard deviation score (SDS) (revised UK reference) in the first and second years of treatment.
Results: Height increased by 0.74 SDS units (SD 0.39) in the first year of treatment and 0.37 units (SD 0.27) in the second. Adjusting for age, height, weight, midparent height, and injection frequency, the strongest predictor of first year growth response was the GH provocation test result; halving the result predicted an extra height increment of 0.09 units ($p<0.0001$). It predicted the second year response less well ($p<0.0002$) and after adjusting for the first year response was not predictive at all.
Conclusions: Among patients referred for possible GHD, the GH provocation test, though not a gold standard for diagnosis, is a valuable predictor of growth response in the first year of treatment. A year's treatment is recommended for cases with a marginal provocation test result, with the option to continue treatment if the response is adequate. The value of unified protocols for single or repeated provocation tests needs to be assessed.

Growth hormone deficiency (GHD) is suspected in subjects with short stature and a reduced growth velocity in whom other causes of poor growth have been excluded.[1] The diagnosis is confirmed by a provocation test which identifies a subnormal response to a GH secretagogue such as insulin, clonidine, glucagon, arginine, or L-dopa. This is generally viewed as more effective than measuring the spontaneous GH secretion[2] though a recent small study suggests the opposite.[3] In subjects with documented hypothalamic/pituitary pathology or additional pituitary hormone deficits, and in those with a classical phenotype and severe isolated GHD (GH response <10 mU/l or <5 ng/ml), the diagnosis is generally straightforward. However the diagnosis of partial GHD (10 mU/l ⩽ GH response <20 mU/l) is more problematic. The various GH provocation tests, and laboratory measures of GH, vary in specificity and sensitivity and the definition of partial GHD is dependent on an arbitrary cut off. A second test may improve the discrimination, but consecutive tests on the same day may be affected by down regulation of the hypothalamic/pituitary axis and require complex statistical analysis.[4] Surrogate measures of GH secretion, such as IGF-1 and IGFBP-3, are useful to rule out the diagnosis but are not sufficiently sensitive for the diagnosis of GHD.[5 6]

The growth response to a standard replacement dose of GH should discriminate those with GHD from those with normal variant short stature, but this response may be influenced by other factors such as midparent height, birth weight, and current weight.[7] Both pretreatment growth rate[8] and stature[9] have been described as predictors of growth response although neither study adjusted for regression to the mean.[10] The peak GH response to pharmacological testing[1 8] and 24 h GH secretion[11] have been shown to be related to growth response. Most recently Ranke and colleagues[12] have developed prediction models for each of the first 4 years of GH treatment, including all the above variables. It is unclear though how useful these models are for deciding whether or not to provide GH treatment.

In order to determine the degree to which the GH provocation test result predicts the response to GH, and to improve the management of patients whose provocation test result is borderline, we analysed data from GH deficient subjects (GH response <20 mU/l) registered in the UK Pharmacia KIGS database.

METHODS

Cases were selected from the UK Pharmacia KIGS (Kabi International Growth Study) database[12] on the basis of a diagnosis of isolated GHD in children <10 years of age treated with GH, Tanner pubic hair stage 1, GH provocation test result <20 mU/l, and follow up >1 year. Cases with a history of pituitary or hypothalamic tumours or cranial radiotherapy were excluded. A total of 337 patients were identified with height measurements at baseline (t0) when GH was first given, and 1 year later (t1). Some patients also had measurements 1 year before (t−1) and 2 years after (t2) baseline. The times were defined to be the nominal time (that is, −1, 1, and 2 years relative to t0) ±2 months. On 5% of occasions patients had more than one measurement in the ±2 month window, which were then averaged.

The KIGS was set up by Kabi Pharmacia (now Pfizer) as a post-market surveillance programme for patients receiving GH. Participating patients (or their parents) gave signed informed consent for their data to be included in the database and used for the purposes of medical research. Irreversibly anonymised data for the current study were provided to the authors by Dr Patrick Wilton (Pharmacia).

The primary outcome measure was the first year growth response, that is, the change in height standard deviation

Abbreviations: GH, growth hormone; GHD, growth hormone deficiency; KIGS, Kabi International Growth Study; SDS, standard deviation score

score (SDS) (British revised reference[13 14]) during the first year of GH treatment, adjusted for other covariates. A secondary outcome was the growth response in the second year of treatment. Growth response was adjusted for the exact time between measurements. Other covariates included height SDS and weight SDS at the start of treatment, age, sex, birthweight SDS, gestation, midparent height SDS (mean of two parents' SDS), GH dose and injection rate, growth rate in the pre-treatment year, and the GH provocation test result. This latter variable was obtained using local protocols including local GH assays. To account for skewness it was logged and multiplied by 100, so that its regression coefficient indicates the effect on height SDS change of a 1% change in the provocation test result.[15]

The factors predicting growth response were identified by multiple regression analysis. The relationship between growth response and provocation test was explored using a partial regression plot, which is a scatterplot where each factor is replaced by its residual adjusted for other factors in the relevant model. Analysis was carried out in Data Desk version 6.2 (Data Description, Ithaca, NY, USA).

RESULTS

Background information

Table 1 summarises the 337 patients at baseline (t0), and during the 2 years of treatment (t1 and t2). Overall there was a preponderance of boys, with midparent height SDS, birth weight SDS, and weight SDS all relatively low. The provocation test result (maximum GH peak) was also low, while details of GH treatment were similar in years 1 and 2.

Height SDS was fairly constant during the year before treatment (mean −3.4 SDS units), increasing to −2.7 after 1 year of treatment and −2.3 after 2 years (table 2). Height SDS at baseline depended on age (0.07 units greater per year, $p = 0.03$), sex (girls 0.3 units shorter than boys, $p = 0.004$), and the GH provocation test result (0.0021 units greater per %, $p<0.0001$).

There was a big growth response in the first year of treatment and a smaller response in the second, which was similar in the two sexes. The SD of height increment was greater in the first year than before or after, indicating the heterogeneity of the growth response.

First year growth response

Table 3 summarises the regression analysis for the first year of treatment. The 10 factors together explained 42% of the variance, with a residual SD of 0.29 height SDS units. The trend in height SDS increment with age was curvilinear, becoming less steep with age (fig 1). Baseline height SDS and weight SDS showed that shorter and heavier children gained more height with treatment, while midparent height SDS predicted a greater response in those with taller parents. The frequency of GH injections was important, but not the total GH dose. Injection frequency was three to seven injections per week, with 28% receiving three, and 79% six or seven.

The GH provocation test result was the most predictive factor, accounting for 9.9% of the variance. Figure 2 shows the partial regression plot between growth response and provocation test, where each factor has been adjusted for the other factors in table 3. The fitted regression line is also shown.

There were also two marginally significant interaction terms (not shown). The provocation test result was less negative (that is, less predictive) in older children (interaction $p = 0.06$), and the GH dose effect was smaller in tall children ($p = 0.09$). There were no significant sex interactions ($p>0.1$), so the sexes were pooled for analysis.

Second year growth response

Just three factors predicted growth response in the second year: provocation test ($p = 0.0005$), GH dose, and injection rate ($p\sim0.05$). The provocation test coefficient was half that seen in the first year. Together the factors explained 6.8% of the variance, and the residual SD was 0.26 height SDS units based on 232 patients.

By far the best predictor of second year growth response was the first year response. Adding it to the above model (along with baseline height SDS at year 1) tripled the explained variance to 21.6%, reduced the residual SD to 0.24 and rendered the provocation test result insignificant ($p = 0.6$) (table 4).

Growth prediction before GH treatment

The one factor predicting growth in the pre-treatment year was age, with older patients growing faster ($p = 0.001$). The

Table 1 Summary statistics at baseline and during years 1 and 2 of treatment in 337 pre-pubertal IGHD patients aged under 10 years (70% male)

Variable	n	Mean	SD
Baseline (t0)			
Age (years)	337	7.8	1.3
Midparent height SDS	324	−1.0	0.9
Birth weight SDS	311	−0.7	1.2
Gestational age (weeks)	316	39	2.5
Maximum GH peak (IU)	337	5.7	4.5
Weight SDS	337	−2.8	1.5
Body mass index SDS	337	−0.4	1.3
Year 1 of treatment			
GH dose (IU/m^2/week)	337	12.5	2.9
GH injections (number/week)	337	5.5	1.7
Year 2 of treatment			
GH dose (IU/m^2/week)	331	13.3	4.1
GH injections (number/week)	322	5.7	1.6

Table 2 Summary statistics of height SDS and change in height SDS for different time periods

Time period	n	Mean	SD	Min	Max
Height SDS at					
t−1	153	−3.4	0.8	−6.4	−1.2
t0	337	−3.4	0.9	−6.4	+0.4
t1	337	−2.7	0.9	−5.1	+1.6
t2	242	−2.3	0.9	−4.9	+1.7
Height SDS change from					
t−1 to t0	153	−0.05	0.21	−0.6	+0.6
t0 to t1	337	0.74	0.39	−0.3	+2.2
t1 to t2	242	0.37	0.27	−0.4	+1.3

Table 3 Multiple regression analysis of the increase in height SDS during the first year of GH treatment on a series of prognostic factors (n = 300)

Variable	Regr. coeff.	SE	t ratio	p
Time between visits (years)	0.63	0.24	2.6	0.009
Age at baseline (years)	−0.52	0.16	−3.3	0.001
Age^2 at baseline (year^2)	0.030	0.010	3.0	0.003
Height SDS at baseline	−0.17	0.03	−5.7	⩽0.0001
Weight SDS at baseline	0.042	0.016	2.6	0.01
Midparent height SDS	0.10	0.02	4.7	⩽0.0001
Birth weight SDS	0.025	0.015	1.6	0.1
Maximum GH peak (%)	−0.0015	0.0002	−7.1	⩽0.0001
GH dose (IU/m^2/week)	0.010	0.006	1.6	0.1
GH injections (n/week)	0.045	0.011	4.2	⩽0.0001

Regr. coeff., Regression coefficient; SE, standard error.

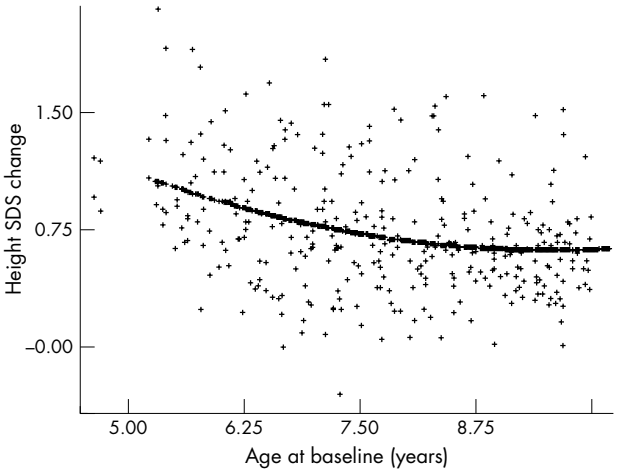

Figure 1 The curvilinear relation between height SDS change and age in the first year of GH treatment.

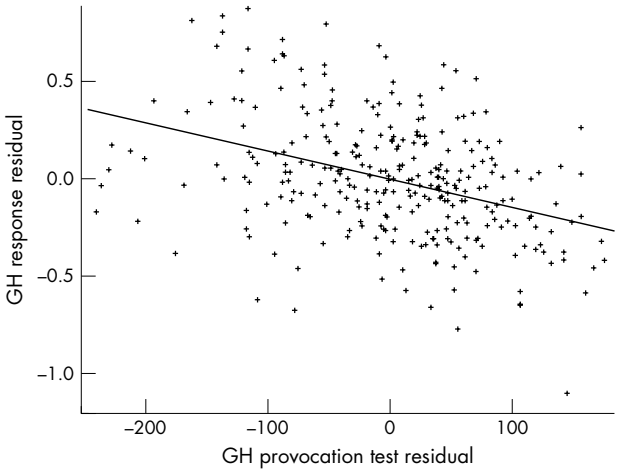

Figure 2 The relation between height SDS change in the first year of GH treatment and maximum GH peak (%), each expressed as residuals adjusted for the other factors in the model of table 3.

analysis was based on 153 patients and explained 6.9% of the variance with a residual SD of 0.20.

Height and height increment correlations

Table 5 gives the correlations between height and height increment at the various time points. Heights prior to GH treatment at t−1 and t0 were highly correlated, $r = 0.97$ indicating strong tracking. Height also tracked strongly during GH treatment, with a correlation of 0.95 between heights at times t1 and t2. The introduction of GH therapy disturbed the tracking, with strikingly lower correlations for heights before and after treatment, for example, 0.9 for the two pre-treatment years versus year 1, and 0.8 versus year 2.

DISCUSSION

This group of very short children, mean height SDS −3.4 at baseline, showed a dramatic increase of 0.74 SDS in height after a year of GH treatment. Despite this their growth pre-treatment was sufficient to maintain their centile position, with a fall of only 0.05 in mean height SDS over the year (table 2). This is at first sight surprising, as untreated GH deficient patients cross height centiles downwards. It suggests the likely inclusion of non-GHD patients in the cohort some of whom were crossing centiles upwards.

A similar conclusion can be drawn from the patients' heterogeneous response to GH treatment. Some were genuinely deficient and grew better with treatment, while others were not deficient and grew no better when treated. This shows itself in several ways. Firstly, there was no association between the provocation test result and growth prior to GH treatment, but a strongly significant and negative association after treatment (table 3). Ranke *et al*[12] also found the provocation test result to be strongly predictive of post-treatment acceleration.

Secondly, the SD of the growth response almost doubled as a result of GH treatment, from 0.21 the year before to 0.39 the year after (table 2). Some patients responded to GH while others did not, and the extra heterogeneity in response on introducing GH was most likely due to the GH itself. The smaller SD of growth response in the second year (0.27) accords with this, in that the GH effect waned after the first year.

Thirdly, before-treatment and after-treatment height tracked strongly (table 5), and almost as strongly as seen in normal children,[16] but heights before and after treatment correlated with each other far less strongly, for example, for the pre-treatment years versus years 1 and 2. This again shows that the introduction of GH treatment generated a heterogeneous response, with some patients responding and others not.

The correlation of 0.45 between height SDS change in the first and second years of treatment (table 5) was remarkably high, given the measurement errors involved. Of course measurement error impacts on all the correlations in table 5, but particularly so for correlations between successive velocities where error in the common measurement (that is, height at t1) counts twice. The high correlation indicates that first year growth response was a good predictor of later response. Furthermore, after adjusting for it the provocation test taken 1 year earlier was completely uninformative (table 4). So the first year response, once known, is far better than the provocation test for predicting final height increment.

The relationship between test result and subsequent growth while treated with GH was essentially linear on the log scale (fig 2). A halving of the provocation test result (for example, from 10 to 5 mU/l) corresponds to a 69% decrease on the log (%) scale,[15] which in turn relates to an increase of $-69 \times -0.0013 = +0.09$ height SDS units, an upward shift in velocity of 0.3 SDs given the residual SD of 0.29 SDS units. This looks a modest velocity effect for a large change in the test result, but if maintained over time it would appreciably increase final height.

The test's relatively poor performance may reflect measurement error due to the lack of standardisation in methodology and inter-laboratory differences. Bright *et al*[17] found that the provocation test had 82% sensitivity but only 25% specificity to detect GHD, low predictive power which they attributed to inter-individual differences in GH handling. It is likely that

Table 4 Multiple regression analysis of the increase in height SDS during the second year of treatment on a series of prognostic factors (n = 232)

Variable	Regr. coeff.	SE	t ratio	p
Maximum GH peak (%)	−0.0001	0.0002	−0.6	0.6
GH dose (IU/m^2/week)	0.014	0.004	3.1	0.002
GH injections (n/week)	0.018	0.011	1.6	0.1
Growth response in year 1	0.31	0.05	6.6	⩽0.0001
Height SDS at year 1	−0.047	0.021	−2.3	0.02

The first year growth response is far more predictive than the provocation test. Regr. coeff., Regression coefficient; SE, standard error.

Table 5 Correlations between height SDS and annual change in height SDS

	Ht SDS at t−1	Ht SDS at t0	Ht SDS at t1	Ht SDS at t2	Change t−1 to t0	Change t0 to t1
Height SDS at t0	**0.97**	1				
Height SDS at t1	**0.88**	**0.90**	1			
Height SDS at t2	**0.81**	**0.80**	**0.95**	1		
SDS change t−1 to t0	−0.11	0.14	0.06	−0.01	1	
SDS change t0 to t1	−0.23	**−0.25**	**0.20**	**0.30**	−0.19	1
SDS change t1 to t2	−0.02	−0.16	0.04	**0.34**	−0.02	**0.45**

Significant correlations are shown in bold ($p<0.001$). n varies between 111 and 337.

both factors are relevant, but using a standardised protocol would be a relatively simple way to improve the signal to noise ratio.

The younger children grew more rapidly than the older children (table 3 and fig 1), and the provocation test result increased with age ($r = +0.20$, $p = 0.0002$). So younger children were more likely to have GHD than constitutional delay, and were more likely to respond to GH treatment. The frequency of GH injections was more significant than the dose during the first year of treatment, but this reversed in the second year. Both the dose and the injection rate were similar in the 2 years, so there is no obvious explanation for the switch. As more and more children receive six to seven injections per week, the dose will become the important factor.

Baseline height SDS had a highly significantly negative effect in the first year of treatment, due partly to regression to the mean and partly to the presence of baseline weight SDS in the model. The inclusion of both weight and height shows that short fat children were those that respond best to GH treatment. This of course describes the classic phenotype of GH deficiency.

Midparent height and birth weight were predictive during the first year of treatment, but not before or after. This suggests that they only emerged as predictive when the growth rate was sufficiently high.

In conclusion, the outcome of the GH provocation test was closely related to growth response in the first year of treatment. Growth response is effectively the ''gold standard'' to quantify the degree of GH insufficiency in the individual patient, since the aim of treatment is to increase height velocity and ultimately adult height. Growth responses in the first 2 years are highly correlated, emphasising the value of first year response for predicting later height gain attributable to treatment.

So the initial GH treatment decision should be based on the provocation test using the conservative cut off of 20 mU/l. Then if the first year growth response is sufficiently small, logically the treatment should be stopped at that point unless compliance is felt to be poor and can be improved. This decision should ideally be based on growth response adjusted for all the factors in table 3 with the exception of the provocation test result. The provocation test justifies treatment in the first place, but after 1 year it is superseded by the growth response and should not be adjusted for again. In practice of course such adjustments are unrealistic without dedicated software.

The way forward is to improve the predictive power of the GH provocation test, by collecting information on local variations in test procedures and comparing the results with those based on a standardised protocol and centralised GH assay. We propose a clinical trial to test the effectiveness of such a protocol.

ACKNOWLEDGEMENTS

We thank the KIGS International Board and Executive Scientific Committee for permission to access the KIGS data, and Dr Patrick Wilton for providing the data. We are also grateful to members of the British Society for Paediatric Endocrinology and Diabetes (BSPED) for contributing their patients to KIGS. Finally we thank the two reviewers for their comments which have considerably improved the paper.

Authors' affiliations

T J Cole, Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, University College, London, UK
P C Hindmarsh, Centre for Human Growth and Maturation, Department of Medicine, University College, London, UK
D B Dunger, Department of Paediatrics, University of Cambridge, Cambridge, UK

Conflicts of interest: all three authors have from time to time received support from Pfizer for contributing to conferences on child growth.

REFERENCES

1 **Milner RD**, Russell-Fraser T, Brook CG, *et al*. Experience with human growth hormone in Great Britain: the report of the MRC Working Party. *Clin Endocrinol (Oxf)* 1979;**11**:15–38.
2 **Rogol AD**, Blethen SL, Sy JP, *et al*. Do growth hormone (GH) serial sampling, insulin-like growth factor-1 (IGF-1) or auxological measurements have an advantage over GH stimulation testing in predicting the linear growth response to GH therapy? *Clin Endocrinol* 2003;**58**:229–37.
3 **Radetti G**, Buzi F, Cassar W, *et al*. Growth hormone secretory pattern and response to treatment in children with short stature followed to adult height. *Clin Endocrinol* 2003;**59**:27–33.
4 **Hindmarsh PC**, Cole TJ. Evidence-based growth hormone therapy prediction models. *J Pediatr Endocrinol Metab* 2000;**13**:1359–64.
5 **Tillmann V**, Buckler JM, Kibirige MS, *et al*. Biochemical tests in the diagnosis of childhood growth hormone deficiency. *J Clin Endocrinol Metab* 1997;**82**:531–5.
6 **Mitchell H**, Dattani MT, Nanduri V, *et al*. Failure of IGF-I and IGFBP-3 to diagnose growth hormone insufficiency. *Arch Dis Child* 1999;**80**:443–7.
7 **Ranke MB**, Gilbaud O. Growth response in prepubertal children with idiopathic growth hormone deficiency during the first year of treatment with human growth hormone. Analysis of the Kabi International Growth Study. *Acta Paediatr Scand* 1990;**370**(suppl):122–30.
8 **Tanner JM**, Whitehouse RH, Hughes PC, *et al*. Effect of human growth hormone treatment for 1 to 7 years on growth of 100 children, with growth hormone deficiency, low birthweight, inherited smallness, Turner's syndrome, and other complaints. *Arch Dis Child* 1971;**46**:745–82.
9 **Wit JM**, Faber JA, Van den Brande JL. Growth response to human growth hormone treatment in children with partial and total growth hormone deficiency. *Acta Paediatr Scand* 1986;**75**:767–73.
10 **Darendeliler F**, Hindmarsh PC, Brook CG. Dose-response curves for treatment with biosynthetic human growth hormone. *J Endocrinol* 1990;**125**:311–6.
11 **Hindmarsh PC**, Smith PJ, Pringle PJ, *et al*. The relationship between the response to growth hormone therapy and pre-treatment growth hormone secretory status. *Clin Endocrinol (Oxf)* 1988;**28**:559–63.
12 **Ranke MB**, Lindberg A, Chatelain P, *et al*. Derivation and validation of a mathematical model for predicting the response to exogenous recombinant human growth hormone (GH) in prepubertal children with idiopathic GH deficiency. *J Clin Endocrinol Metab* 1999;**84**:1174–83.
13 **Freeman JV**, Cole TJ, Chinn S, Jones PRM, *et al*. Cross-sectional stature and weight reference curves for the UK, 1990. *Arch Dis Child* 1995;**73**:17–24.
14 **Preece MA**, Freeman JV, Cole TJ. Sex differences in weight in infancy [letter]. *BMJ* 1996;**313**:1486.
15 **Cole TJ**. Sympercents: symmetric percentage differences on the 100 $\log_e$ scale simplify the presentation of log transformed data. *Stat Med* 2000;**19**:3109–25.
16 **Bailey BJR**. Monitoring the heights of prepubertal children by the use of standard charts. *Ann Hum Biol* 1994;**21**:1–11.
17 **Bright GM**, Julius JR, Lima J, *et al*. Growth hormone stimulation test results as predictors of recombinant human growth hormone treatment outcomes: preliminary analysis of the National Cooperative Growth Study Database. *Pediatrics* 1999;**104**:1028–31.