Available online at www.sciencedirect.com
SCIENCE DIRECT®
Growth Hormone & IGF Research 14 (2004) 185–194
www.elsevier.com/locate/ghir

Review

Is growth hormone stimulation testing in children still appropriate?

Laura M. Gandrud *, Darrell M. Wilson

Division of Pediatric Endocrinology and Diabetes, Stanford University Medical Center, S-302, Stanford University, Stanford, CA 94305-5208, USA

Received 30 July 2003; received in revised form 15 November 2003; accepted 18 November 2003

Abstract

The diagnosis of growth hormone deficiency (GHD) historically has relied on measurement of growth hormone (GH) concentrations following stimulation, usually with a non-physiologic provocative agent. Despite the use of more specific GH assays, the peak concentration of GH below which a child is considered GH deficient has risen. We examine the pitfalls associated with GH stimulation tests, specifically, the lack of reliability and accuracy of these tests, and their inability to predict who will benefit from GH therapy. We recommend that GH stimulation tests no longer routinely be used for the diagnosis of GHD in children.

Keywords: Growth hormone deficiency; Insulin tolerance test

1. Introduction

Documentation of growth hormone deficiency (GHD) has historically relied on testing growth hormone (GH) levels after stimulation of its release, or in some centers, frequent sampling of GH levels. Clinical features, additional biochemical measures, and imaging often contributed to the decision to perform these tests. When first available in the 1960s, supplies of pituitary derived GH were extremely limited. To ensure that this scarce resource was appropriately rationed to those most likely to respond, GHD had to be confirmed with a stimulated GH concentration of less than 3 ng/mL.

Following the introduction of essentially unlimited supplies of synthetic GH in the 1980s, the use of GH has exploded. Paradoxically, as GH assays grew more specific, criterion for failing a GH stimulation test grew more generous. Cutoffs rose from 3 to 7 ng/mL to the current 10 ng/mL level.

The number of GH stimulation protocols has also increased. An exhaustive review by Sizonenko et al. [1] examined the wide variation in stimulation testing: "at least 34 provocative tests have been developed and 189 different combination protocols were found. Most investigators have used these tests without normative values for age, pubertal status or gender". There is no consensus as to what is the definitive stimulation test for GHD.

Although GH is now widely available, the accurate diagnosis of GHD remains important for the diagnosis of underlying disorders and for treatment decisions. Diagnosis and treatment of GHD in infants can prevent hypoglycemia. Evaluating the cause of GHD can lead to the diagnosis of congenital anomalies, other hypothalamic–pituitary abnormalities, genetic mutations or acquired pathology. The Food and Drug Administration approved the use of GH for the treatment of children with idiopathic short stature (ISS). Despite FDA approval, the benefits of GH therapy in this group of children remain controversial. Response to GH therapy in children with GHD is characteristically excellent; therapy has been shown to be more cost-effective in children with GHD as compared to ISS [2]. Fortunately, a wide array of diagnostic procedures, not available when GH stimulation testing was first established aid in the diagnosis of GHD in children.

Is GH stimulation testing still appropriate for the diagnosis of GHD in children? We believe that it is not.

* Corresponding author. Tel.: +1-650-723-5791; fax: +1-650-725-8375.
E-mail address: Lgandrud@stanford.edu (L.M. Gandrud).

1096-6374/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ghir.2003.11.003

A good medical test should be precise, accurate and concordant with the diagnosis or outcome [3]. A good medical test is one whose outcome is used to guide therapy and to predict response to therapy. Among testing protocols of similar efficiency, the test that is safe, easy and inexpensive is preferred. We review the evidence that GH stimulation testing does not meet these criteria. Provocative testing often misdiagnoses normal children as GH deficient. GH stimulation testing may occasionally misdiagnose children who are truly GH deficient as GH sufficient, creating obstacles for timely and appropriate GH therapy as well as insurance coverage.

We review the quality, precision, accuracy and usefulness of GH stimulation tests in clinical management. We present data demonstrating the need for alternatives to GH stimulation testing in the diagnosis of GHD, and data that suggest it is inappropriate to continue to use these GH stimulation tests.

2. Is GH stimulation testing precise?

2.1. Short-term precision

Precision refers to the reproducibility of a test. When the test is repeated in the same patient, on multiple occasions, the results should be similar. There have been a limited number of studies looking at the short-term precision of GH stimulation tests. Tassoni et al. [4] performed repetitive GH stimulation tests in 49 children with an average height standard deviation score (SDS) of −2.34 (range from −3.79 to −1.85) and a height velocity between the third and 25th percentiles. The children (mean age, 11.2 years) underwent arginine stimulation and sleep tests, or arginine and L-DOPA stimulation tests. The studies were repeated one to three weeks later. The mean peak GH response to arginine was 6.9 ± 5.5 ng/mL (mean ± SD). The coefficient of variation for arginine ((the difference between study results/mean of the two studies) × 100) was extremely high, 89% in one group, and 66% in the other, indicating poor reproducibility. The coefficient of variation for L-DOPA was 86%. Using a cutoff of 7 ng/mL, 21% of patients passed each arginine stimulation test, 37% failed each test, and 42% had discordant results. Hoeck et al. [5] performed repeat insulin tolerance tests in 16 non-obese healthy adult subjects. The testing was performed at least 72 h apart. No statistical correlation was found between the results of the first test as compared with the repeat test (Fig. 1). Zadik et al. [6] similarly showed limited short-term reproducibility with a correlation between consecutive GH stimulation tests of 0.52.

Carel et al. [7] studied the GH stimulation test results of 3233 children who had received GH in France between 1973 and 1989. A total of 62 different pairs of stimulation tests were used. The time between the two studies was not published. The mean intraclass correlation coefficient value was just over 0.5 for all combinations of stimulation tests studied.

2.2. Long-term precision

Repeat provocative testing after long-term GH therapy *repeatedly* fails to confirm the childhood diagnosis of GHD. Cacciari et al. [8] retested 184 children (113 male, 71 female) who had been diagnosed as GHD by having both a GH response less than 8 ng/mL to L-DOPA and arginine (Test #1), and a mean overnight GH concentration less than or equal to 3.3 ng/mL. The children were retested 2.8 years after the start of GH therapy (Test #2). On retesting, 18.5% of the subjects who were initially defined as GHD (i.e., GH concentration of less than 8 ng/mL) had normal GH secretion

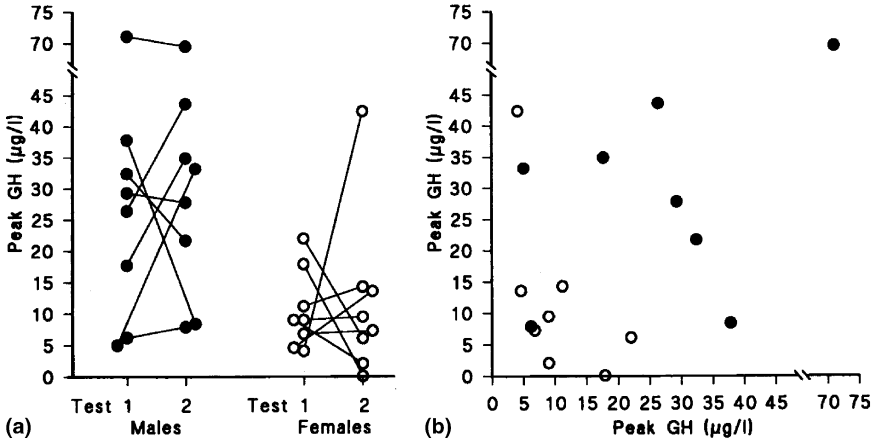

Fig. 1. Serial GH stimulation testing. Responses to the first and second GH stimulation test in each individual are shown in (a). In (b), the subject's peak GH response in test 1 (*X*-axis) is plotted against the peak GH response in test 2 (*Y*-axis). Open circles represent females; closed circles, males [5]. (Reprinted with permission from the Society of the European Journal of Endocrinology.)

on both stimulation and overnight testing. 68 of 184 children were again retested, 1.5 years later (Test #3). Nine (14.5%) subjects normalized on Test #3, and 45.6% had results that were discordant from Test #2. All six subjects who had normalized on Test #2 did not have normal GH secretion on Test #3.

Maghnie et al. [9] looked at the response to retesting after GH therapy based on previous brain images. All 18 patients who had a normal or small pituitary on initial magnetic resonance imaging (MRI) had a normal response following the second arginine/insulin tolerance test. In contrast, all 17 subjects who had either pituitary hypoplasia ($n = 13$) or craniopharyngioma ($n = 4$) had consistent GH responses of <3 ng/mL when retested. The authors propose that only those patients with abnormal MRI findings had permanent GHD. They recommend that children with normal or small pituitary glands should undergo retesting long before attainment of final adult height, so that GH therapy could be discontinued. Coutant et al. [10] found that on retesting after GH therapy, 22 of 35 (63%) subjects with normal MRI findings had a peak response >10 ng/mL, while none of seven subjects with abnormal MRI findings had normalized on retesting.

Interval GH therapy and/or puberty might contribute to the normalization of responses in children initially found to be GH deficient by GH stimulation testing. To address this hypothesis, Loche et al. [11] performed serial GH stimulation tests in subjects who were neither treated with GH nor entered puberty in the interval between testing. Thirty-three prepubertal children (5.2–10 year) who had initially failed two GH stimulation tests (clonidine, arginine, or insulin tolerance) with responses on both of <10 ng/mL, underwent repeat testing one to 6 months later. All subjects had a normal MRI and short stature (< −2SD) and/or poor growth velocity (<25%). GH response in the second stimulation test was greater than 10 ng/mL in 28 of 33 (85%) subjects (Fig. 2). Of nine patients with initial responses less than 7 ng/mL, eight had a response on repeat testing of >10 ng/mL while one subject had a peak response of 9 ng/mL. They found no factors predictive of response on retesting. The authors suggested repeat testing of children with normal MRI who have initially poor response to provocative tests. Having removed the effects of puberty and GH therapy, the inherent poor reproducibility of the stimulation tests is undeniable.

GH stimulation tests are not precise; results are not reproducible after a short or long period of time across a wide range of patients. A number of factors may be involved in the poor precision. The pituitary response can be affected by previous stimuli as well as somatostatin levels. The peak response could be missed as a result of sampling error. There may be a longer-than-yearly cyclic or seasonal variability in GH secretion [12]. Retesting also results in regression to the mean, best described by van den Broeck et al. [13]. Whether there is an entity of transient GHD that may account for discordant results over a long-term period remains controversial [11,13–15]. Transient GHD is difficult to prove given the poor reproducibility of the stimulation tests [13].

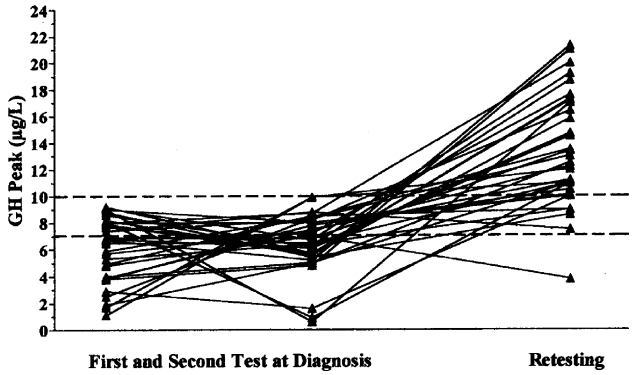

Fig. 2. Serial GH stimulation testing [11]. Each subject's peak GH response to stimulation with insulin, clonidine, or arginine is marked by a triangle. The first and second GH stimulation tests occurred at the time of diagnosis; retesting occurred one to 6 months after diagnosis. Solid lines connect the responses of each subject. The dotted lines represent the cutoffs of 7 and 10 ng/mL. (Reprinted with permission from the Journal of Pediatrics.)

3. Is GH stimulation testing accurate?

Accuracy here refers to a test's agreement with a diagnosis, as determined by other measures or standards (often referred to as a gold standard). Given the problems with precision noted above, the insulin tolerance test using the current cutoff for failure should not be considered the gold standard for the diagnosis of GHD. An accurate test differentiates the diseased from the disease-free, i.e., those with GHD from those with normal GH secretion. *Without good precision, accuracy is essentially impossible.*

3.1. Gold standard

For the diagnosis of GHD, the gold standard has been difficult to define. The circular reasoning of using the test under study as the gold standard must be avoided. Thus, one should not diagnose GHD with a GH stimulation test (such as the insulin tolerance test, often considered the gold standard) and then claim that the GH stimulation test is superior to other possible methods for determining GHD, a common error in the literature. Given the imprecision of GH stimulation tests discussed above, the lack of concordance of alternative tests to diagnose GHD with the results of GH stimulation tests is hardly surprising.

In 1968, Goodman et al. [16] originally describe the clinical and biochemical features of children with GHD. The 16 children with isolated GHD had poor growth velocity (from −4.0 to −8.2 SDS), and a majority presented with growth failure by 2 years of age. The children had a pudgy appearance with truncal obesity and round facies. Symptomatic hypoglycemia was common. GH therapy resulted in a significant improvement in growth velocity. These children had a mean peak response to GH stimulation tests (insulin tolerance test or arginine) of 1.3 ng/mL with *a maximum response of 2.5 ng/mL*. The consistently low responses were similar to those with multiple pituitary deficiencies (mean peak response of 1.2 ng/mL and a maximum response of 2.5 ng/mL) and to those with organic pituitary lesions as described by Kaplan et al. [17].

Tillman et al. [18] studied a group of children diagnosed with GHD based on clinical and auxologic data, along with a group of short children without GHD. The children underwent arginine, glucagon, clonidine, insulin or exercise stimulation tests; some received a sex steroid prime. A subgroup of children with "definitive GHD" (hypothalamic–pituitary pathology such as septo-optic dysplasia, congenital hypopituitarism, tumors, or abnormal findings on computed tomography or MRI) had a mean peak response of 3.9 ng/mL with a range from 0.5 to 13.0 ng/mL. When looking at all children with a diagnosis of GHD, sensitivity of the GH stimulation test using a cutoff of 7.5 ng/mL was 73% with a specificity of 85%. The positive predictive value at this cutoff was only 50%. In contrast, at a cutoff of 2.5 ng/mL, the sensitivity was 53%, specificity, 92% and positive predictive value, 80%.

Adan et al. [19] studied subgroups of children who had been defined as GH deficient based on peak GH response of <10 ng/mL to two stimulation tests. Those with "certain GHD" (abnormal hypothalamus–pituitary on imaging, microphallus or hypoglycemia) had a mean GH peak response to arginine/insulin or glucagon of 3.9 ± 0.6 ng/mL. Children with "transient GHD" (with history of a normal response to stimulation while GH therapy was suspended) had a mean peak response of 7.7 ± 0.5 ng/mL and those with "uncertain GHD" had a mean peak response of 6.7 ± 0.8 ng/mL. The differences between the transient and "uncertain" GHD groups were not statistically significant.

GH hormone stimulation tests, with the limitations discussed, perform poorly when compared to other means of diagnosing GHD, such as a combination of auxology and cranial imaging.

3.2. Cutoff for abnormal response

The cutoff used to determine when a test is abnormal influences the accuracy of the test. Which peak GH concentration cutoff during GH stimulation testing best differentiates those with GHD from those with normal GH secretion? In the past, 3 ng/mL was the cutoff commonly used. Kaplan et al. [17] studied 134 prepubertal short children and 10 controls. Fifty-three subjects had hypopituitarism (16 isolated, 19 multiple hormone deficiencies, 15 organic lesions, three diabetes insipidus (DI)), the others had "primordial dwarfism", constitutional short stature, or delayed puberty. Each underwent insulin tolerance testing. The mean GH response to provocation was 2.5 ng/mL in 52 of 53 children with hypopituitarism, 12.4 ng/mL in 10 control subjects, 12.5 ng/mL in 20 children with constitutional short stature, 11.8 ng/mL in five children with delayed puberty, and 7.0 ng/mL in eleven children with miscellaneous disorders (renal/gastroenterologic anomalies, Cushing syndrome, chronic disease, malabsorption, etc.). The difference between those with hypopituitarism and the other disorders was both statistically and clinically significant, and emphasizes the validity of a cutoff substantially lower than the 10 ng/mL commonly used today. Similar results have been demonstrated in adults, in whom a cutoff of 3 ng/mL is still routinely used for diagnosis [20].

Youlton et al. [21] confirmed the usefulness of the <3 ng/mL cutoff. Sixty children with growth retardation underwent arginine/insulin tolerance testing. The children were placed in three groups depending on the peak GH response. Group I consisted of 19 children with a peak GH of greater than 7 ng/mL (mean, 14 ng/mL) following both arginine and insulin testing; no other pituitary hormone deficiencies were noted. Group II consisted of 18 children with a response of $\leqslant 3$ ng/mL on both tests. Group II's minimal GH response to stimulation was consistent with their clinical findings. Twelve of 18 had physical stigmata of GHD, while the remaining six had DI (secondary to craniopharyngioma ($n = 3$), histiocytosis ($n = 1$), teratoma ($n = 1$), and undetectable intracranial lesion ($n = 1$)). Group III consisted of 23 children with discordant results, with a peak GH of <7ng/mL after only one of the GH tests. Only one child in Group III group had clinical characteristics of GHD; and no children in this group had multiple pituitary hormone deficiencies.

Today, approximately 70% of pediatric endocrinologists use a cutoff of 10 ng/mL [22]. This arbitrary cutoff has undermined the already limited utility of the test, given the unacceptable number of false positive results. The high false positive rate of the tests is revealed when children without GHD, and with normal growth velocity, undergo GH stimulation testing. This has been evident from Kaplan's early studies in 1968 [17]. The range of peak GH response to stimulation in controls with normal stature was 3.4–28.4 ng/mL; three of 10 children had a maximum response less than 10 ng/mL. The mean peak GH response in children with constitutional shortness of stature was 12.5 ng/mL with a range

of responses of 3.2–24.3 ng/mL. Eight of 20 children had a response less than 10 ng/mL.

These findings have been reproduced in subsequent studies. Marin et al. [23] studied the GH response to provocative stimuli in 84 normal children with heights between the 2.5th and 97.5th percentile and normal growth velocities (Fig. 3). The children underwent three GH stimulation tests, exercise (treadmill), arginine, and insulin, and the peak value from these three tests was analyzed. The mean peak GH response was 6.9 ng/mL in Tanner 1 subjects and 9.0 ng/mL in Tanner 2 subjects. Using a cutoff of 10 ng/mL, 75% of normal prepubertal children would have been "diagnosed" with GHD. A substantial proportion of normal pubertal children also had peak GH concentrations below the current criteria for GHD.

Ghigo et al. [24] demonstrate that the GH response to a number of provocative stimuli often failed to raise the GH levels above 7–10 ng/mL in 472 children without a diagnosis of GHD and with a normal growth velocity. The minimum response on testing in these children without GHD ranged, depending on the particular stimuli used, from 0.5 to 3.8 ng/mL. Twenty-three to 36% of the subjects receiving arginine, clonidine, L-DOPA or glucagon had a GH response less than 10 ng/mL. The authors conclude that "the finding of such low GH responses in normal subjects makes these tests unsuitable for discriminating between normal and GHD subjects" [24]. A number of other studies involving GH stimulation tests in healthy adults and children with normal growth velocity have demonstrated similar rates of false positive results, if the cutoff for poor response is 10 ng/mL [5,6,20,25].

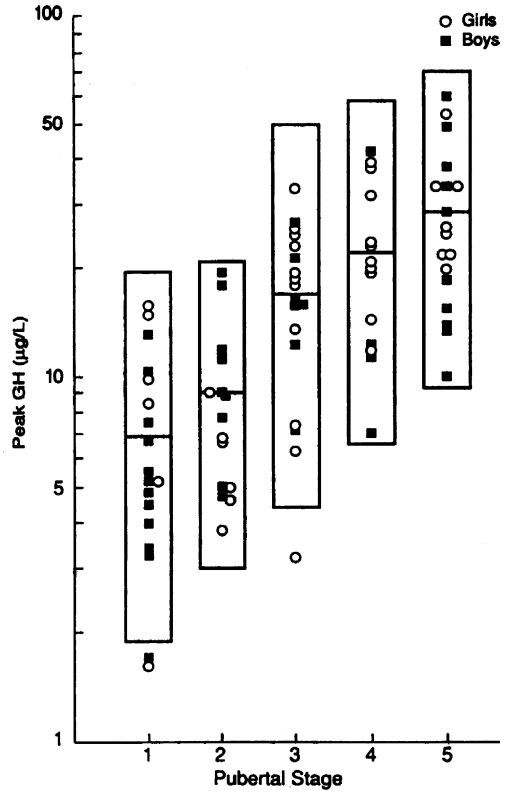

Fig. 3. The peak GH response to GH stimulation with exercise, insulin, or arginine based on pubertal status. Boxes represent the 95% confidence intervals; the line within the box is the mean GH response [23]. (Reprinted with permission from the Endocrine Society.)

Clinicians in the US continue to use a response less than 10 ng/mL for diagnosis of GHD despite the lack of a scientific basis for the arbitrary cutoff. The specificity of GH stimulation tests does improve when a lower cutoff is used [18,24]. If GH stimulation tests are to be used, an appropriately lower cutoff (3–5 ng/mL) for diagnosis of GHD should be used.

3.3. Are there factors influencing the accuracy of GH stimulation tests?

Unlike most other stimulation tests in endocrinology, the arbitrary cutoff defining GH deficiency generally has not varied depending on the provocative agent or GH assay used. Moreover, the pubertal status of the child is not incorporated in the analysis despite evidence that these factors influence test results. Carel et al. [7] reviewed the GH stimulation test results in 3233 children who received GH therapy. A number of factors were isolated that independently correlated with peak GH levels on stimulation testing. Age, puberty, and height SDS were positively correlated with peak GH levels, while weight SDS and genetic target height SDS were negatively correlated. These various factors affecting peak GH response make the use of a single GH concentration as criteria to define GHD untenable.

3.3.1. Growth hormone assays

As the cutoff defining GH deficiency has risen, GH assays have improved such that GH levels using newer assays are two to threefold lower than with older assays [26]. There is also much inter-assay variability. Granada et al. [27] performed overnight GH secretion studies using 11 different GH assays (five radioimmunoassays, five monoclonal immunoradiometric assays, and one monoclonal enzyme assay). The highest result was three times higher than the lowest result.

Mauras et al. [28] studied 13 children with normal stature and 19 children with short stature who underwent GH stimulation testing. The peak GH concentrations between the short stature and normal stature groups were not different. The peak response differed slightly based on the assay used; four of nine controls had a peak response of less than 7 ng/mL using the Hybritech immunoradiometric assay (IRMA), six of 13 using the Diagnostic Systems Laboratories (DSL) ELISA and 11 of 13 using the immunofunctional GH (IFGH) assay. The proportions improved slightly using a cutoff of 5 ng/mL: three of nine with the IRMA, and six of 13 with the

DSL ELISA were below this lower cutoff. There have been other reports which document that different assays result in variable GH levels. After evaluation of six different GH assays performed in four different laboratories, Andersson et al. [29] conclude that "it is apparent that the same cutoff value is not appropriate for interpretation of test results measured by different assays".

Wyatt et al. [22] emphasize the need for standardization of GH assays given that the variability of assays contributes to the lack of accuracy of GH stimulation tests. Standardization would certainly reduce the confusion introduced by the different assays and would improve interpretation of the stimulation tests. Unfortunately, despite pleas to standardize made nearly a decade ago, this has not been done.

3.3.2. Provocative agent

Different provocative agents have been shown to result in discordant results when used in the same patient. Cavallo et al. [30] report that when performing insulin and GH releasing hormone (GHRH) tests in "normal short children", 19 of 53 (36%) were discordant, with 12% (9/77) discordance when using clonidine and GHRH. Ghigo et al. [24] studied the GH response to a number of different stimuli in 472 children with normal stature and 177 with "normal short stature". Mean GH response to physical exercise, insulin-induced hypoglycemia, arginine, clonidine, L-DOPA, glucagon, pyridostigmine (PD), GHRH, PD/GHRH, and arginine/GHRH was 12.7–61.8 ng/mL, with the largest response resulting from arginine/GHRH stimulation.

Tzanela et al. [31] demonstrate the impact that somatostatin tone can have on GH response to stimulation. Using a combination of GHRH and somatostatin, the GH stimulation tests discriminated between children with GHD (defined by previous peak response to stimulation of less than 7 ng/mL) and those with ISS (peak response greater than 10 ng/mL). How to interpret the accuracy of these stimulation tests, when compared to the results of other stimulation tests and not to an independent gold standard, is unclear. Despite the superiority of these combinations of provocative agents, 189 combination protocols are still in use. The discordant response to different stimuli invalidates a single universal cutoff for GH stimulation tests.

3.3.3. Pubertal status/sex steroid prime

Whether the pubertal status of a child or whether the use of a sex steroid prime impacts the peak GH response to stimulation remains controversial. Cavallo et al. [30] report that in 574 short normal children and adolescents, pubertal stage did not affect GH response to stimulation, although the study did not include subjects with Tanner 4 or 5 pubertal development. Similarly, Ghigo et al. [24] did not find a difference in response based on pubertal status.

Other studies demonstrate benefits of sex-steroid priming. In 84 children with normal stature and growth velocity, Marin et al. [23] show an increased peak GH response (highest value after treadmill exercise, arginine, or insulin stimulation) following two days of estrogen priming. Gonc et al. [32] demonstrated a similar increase in peak GH response with use of a testosterone prime. In a double-blinded randomized controlled trial, 15 children with GHD (based on clinical, biochemical, auxologic, and imaging characteristics) and 44 prepubertal or early pubertal healthy short children received estradiol or placebo daily for three days prior to stimulation with arginine/clonidine [33]. Four weeks later, the children received the other form of therapy and underwent a second stimulation test. With estradiol priming, a peak GH response of 9.0 ng/mL (NIH polyclonal GH RIA, equivalent to a level of 3–5 ng/mL on current assays) differentiated the two groups with 100% sensitivity and 98% specificity. Without priming, the sensitivity was 86.7% and the specificity 90.9%, using a cutoff of 5 ng/mL. Despite these data, many clinicians do not prime with estradiol before stimulation nor use the more appropriate cutoff of 3–5 ng/mL.

4. Is growth hormone stimulation testing in neonates necessary?

The diagnosis of GHD in neonates is unique. Miller et al. [34] documented the GH secretory patterns of neonates, studied at an early age (28.2 h after birth) or at a later age (74.8 h after birth). GH release was pulsatile, with decreasing frequency, peak and nadir of pulses in older neonates. The average GH nadir was 13–23 ng/mL, much higher than children and adolescents. But random GH levels are not often drawn within the first 72 h of life and alternatives are necessary. GH levels are often obtained at the time of spontaneous hypoglycemia; a "passing" level is likely around 30 ng/dL [35]. Bhala et al. [36] demonstrate that in this select population, IGFBP-3 is a useful marker of GHD. IGF-BP2 levels are high in the setting of GHD and may have a diagnostic role in infants as well [37]. GH stimulation tests in this population are difficult to interpret in light of the high baseline GH levels and are unnecessary given the available alternatives.

5. Do GH stimulation tests predict response to GH therapy?

Although provocative tests using cutoffs of 7–10 ng/mL fail to accurately distinguish those with true GHD, from those with normal GH secretion, they might have some utility if they were able to predict response to therapy. A preliminary analysis of the National Coop-

erative Growth Study database looked at 236 prepubertal children with a pre-treatment diagnosis of idiopathic GHD or short stature [38]. Children who had a height increase of 0.5 SDS during the first year of GH therapy were defined as having a positive response to therapy. Fifty-six percent of subjects failed the GH stimulation test (<10 ng/mL) and had a positive response to therapy. Eight percent of children passed the GH stimulation test and did not have a positive response to therapy. Overall the GH stimulation test predicted only 64% of the responses to GH therapy, slightly better than a coin toss. The sensitivity of the test was 82%, while specificity was a mere 25%.

Van den Broeck et al. [39] reviewed the Dutch Growth Hormone Register and analyzed whether long-term response to therapy of 435 children could be predicted by peak GH response at the time of diagnosis. Children were diagnosed as severe GHD, partial GHD, or ISS based on GH response to GH stimulation tests (<10 mU/L or 5 ng/mL, 10–20 mU/L or 5–10 ng/mL, >20 mU/L or 10 ng/mL, respectively). The response to therapy was significantly better in children with severe GHD, but was not statistically different between children with partial GHD and those with ISS. The authors conclude that "there is a lack of prognostic validity of GH stimulation tests as they are currently applied" [39].

6. Are GH stimulation tests safe and inexpensive?

GH stimulation tests are not without risk. Use of insulin stimulation testing can result in hypoglycemic seizures. Studies in young children with diabetes demonstrate the negative long-term effects hypoglycemic seizures have on cognitive function [40–42]. Galloway et al. [43] used a strict protocol for insulin tolerance tests in 550 children and reported no severe adverse events. Despite reassuring reports, there have been two reports of death and one report of neurological damage following insulin GH stimulation testing complicated by hypoglycemia [44]. Stimulation with clonidine has also been associated with hypoglycemia [45]. Arginine stimulation testing can result in hypotension and anaphylactoid reactions.

The cost of these studies is not inconsequential. The cost of a GH stimulation test in our facility (Lucile Packard Children's Hospital at Stanford), including administration of arginine and insulin, nurse monitoring and laboratory studies totals over $1000.

7. Do GH stimulation tests impact clinical decisions?

If the results do not guide medical therapy, what purpose does a test serve? Despite the evidence of the low reproducibility and poor accuracy of GH stimulation tests, they continue to be commonly performed. Many groups recommend trials of GH therapy in children with poor growth velocities and normal responses to provocative testing [4,21,46]. The recently published consensus guidelines [47] of the Growth Hormone Research Society and the Lawson Wilkins Pediatric Endocrine Society [58] state that children with repeatedly low insulin-like growth factor-1 (IGF-1) and/or insulin-like growth factor binding protein-3 (IGFBP-3) levels (in the absence of systemic disorder which may affect growth factor production), and normal response on GH stimulation testing, could be considered for GH therapy on the basis of a possible abnormality of the GH/IGF axis. Others advocate postponing therapy of children with an abnormal response to provocative testing if the head MRI is normal, until reevaluation [11]. These recommendations lead to clinical decisions which are contrary to results of the GH stimulation test. Why perform a costly, cumbersome, imprecise and inaccurate test if it does not impact decisions about therapy?

8. How should GHD be diagnosed?

The objective of this paper is not to provide a clinical pathway for the diagnosis of GHD, nor to present all data on alternatives to GH stimulation tests to diagnose GHD, but a few are outlined. Diagnosis of children with GHD can often be made based on the history of growth arrest, pituitary surgery or irradiation, auxology, physical stigmata, and/or the presence of other pituitary hormone deficiencies. These children comprise the group that would have a response of less than 3 ng/mL to stimulation as initially described by Kaplan and Goodman [16,17]. Their diagnosis is clear, and does not require provocative testing to confirm. As Frasier [48] states, "The appropriate focus is on the patient and not on any particular set of numbers". These children must undergo imaging of the pituitary and hypothalamus before initiation of GH therapy. Frequently the diagnosis is not as clear and other diagnostic studies are required. Promising alternatives include growth factor levels, imaging of the brain, and genetic studies.

8.1. IGF-1 and IGFBP-3

IGF-1 and IGFBP-3 have been proposed as tools to diagnose GHD in children [9,11,19,20,28,49–51]. Often, the biochemical data are compared to the results of GH stimulation tests. GH stimulation tests as discussed are a poor gold standard to judge the concordance of other tests with disease. This makes interpretation of these studies difficult. Adan et al. [19] studied the IGF-1 and IGF-BP3 levels in children with "certain GHD" (pituitary stalk interruption, familial, and/or hypoglycemia

and microphallus), "transient GHD" (normal GH peak after a third stimulation test), and "uncertain GHD" (abnormal response to GH stimulation but without MRI abnormality, or hypoglycemia/microphallus) and compared the levels to those in children with ISS. The IGF-1 and IGFBP-3 levels were significantly lower in the children with certain GHD as compared to those with transient or uncertain GHD, as well as children less than 7 years with ISS. In the children with certain and transient GHD, the sensitivity of combined IGF-1 and IGFBP-3 was 96% and the specificity 92%. Elevation of IGFBP-2 is also noted in the setting of GHD.

It has been demonstrated that lower IGF-1 levels prior to initiation of therapy correlates with better responses to therapy. Ranke et al. [52] found that in 156 subjects with GHD (defined by results of GH stimulation testing) and 153 subjects without GHD, the first year growth response to GH correlated with IGF-1 and IGFBP-3 at the start of therapy.

8.2. Magnetic resonance imaging

Improved magnetic resonance imaging has contributed to the diagnosis of GHD [53]. Bozzola et al. [54] demonstrate the association between hypothalamic/pituitary abnormalities and GHD. Coutant et al. [10] demonstrate that MRI findings are more consistent with clinical characteristics of GHD and more predictive of response to GH therapy than are peak GH responses. Subjects with GHD (defined by peak GH on stimulation <10 ng/mL) and abnormal hypothalamic/pituitary on MRI had a significantly better response to therapy than did subjects with GHD and normal MRI findings. The response to GH therapy among children with "complete GHD" (peak GH <5 ng/mL) and "partial GHD" (GH 5–10 ng/mL) was similar if the head MRI was normal. All 15 patients with abnormal MRI findings had multiple pituitary hormone deficiencies, while only one of 48 subjects with "GHD" (defined by GH stimulation testing) and normal MRI findings had multiple pituitary deficiencies. Additionally, the growth response to therapy was not significantly different between children with partial GHD and those with ISS who received therapy. The authors conclude that imaging with MRI is a more effective way to predict those children who are truly GHD and who in turn will have an excellent response to therapy. They suggest that children with a GH peak of <10 ng/mL and a normal MRI actually have other genetic causes of short stature, and that partial GHD may only be one etiologic factor of many. This evidence is in agreement with that of Maghnie et al. [9] who showed that the majority of children with normal MRI who initially had GH peak of <10 ng/mL, subsequently had normalized results on retesting. Zenaty et al. [55] demonstrate the usefulness of head imaging results to predict growth response to 3 years of GH therapy.

8.3. Genetic testing

A number of gene mutations (GH-1, GHRH-1, and PROP-1) resulting in GHD have been determined, allowing for definitive diagnosis of familial GHD in many cases. Genetic mutations resulting in GH resistance, IGF-1 deficiency and abnormalities in the signal transduction pathway are being sought. Interestingly, children with clinical evidence of isolated GHD, but normal MRI findings may represent those with mutations affecting GH secretion [56,57]. It is likely that the number of identified mutations will grow rapidly, increasing the usefulness of genetic testing in the diagnosis of GHD.

9. Conclusion

GH stimulation tests are no longer appropriate for the routine diagnosis of GHD in children. As currently used, these tests lack precision, accuracy, and have little concordance with disease. The tests are poor predictors of response to GH therapy, cumbersome, expensive, and can have side effects. Just as measurement of T_4 and TSH has replaced protein bound iodine as measures of thyroid function and head MRI replaced pneumoencepaholgram, so too should GH stimulation tests be replaced.

While efforts should focus on establishing a more effective algorithm for the diagnosis of GHD and on improving the prediction of response to GH therapy, the time has come to stop doing routine GH stimulation tests in children.

References

[1] P.C. Sizonenko, P.E. Clayton, P. Cohen, R.L. Hintz, T. Tanaka, Z. Laron, Diagnosis and management of growth hormone deficiency in childhood and adolescence. Part 1: diagnosis of growth hormone deficiency, Growth. Horm. IGF Res. 11 (2001) 137–165.

[2] H.J. Guyda, Four decades of growth hormone therapy for short children: what have we achieved?, J. Clin. Endocrinol. Metab. 84 (1999) 4307–4316.

[3] H.C. Kraemer, Evaluating Medical Tests: Objective and Quantitative Guidelines, Sage Publications, Newbury Park, 1992, p. 294.

[4] P. Tassoni, E. Cacciari, M. Cau, et al., Variability of growth hormone response to pharmacological and sleep tests performed twice in short children, J. Clin. Endocrinol. Metab. 71 (1990) 230–234.

[5] H.C. Hoeck, P. Vestergaard, P.E. Jakobsen, P. Laurberg, Test of growth hormone secretion in adults: poor reproducibility of the insulin tolerance test, Eur. J. Endocrinol. 133 (1995) 305–312.

[6] Z. Zadik, S.A. Chalew, Z. Gilula, A.A. Kowarski, Reproducibility of growth hormone testing procedures: a comparison between 24-h integrated concentration and pharmacological stimulation, J. Clin. Endocrinol. Metab. 71 (1990) 1127–1130.

[7] J.C. Carel, J.P. Tresca, M. Letrait, et al., Growth hormone testing for the diagnosis of growth hormone deficiency in childhood: a population register-based study, J. Clin. Endocrinol. Metab. 82 (1997) 2117–2121.

[8] E. Cacciari, P. Tassoni, A. Cicognani, et al., Value and limits of pharmacological and physiological tests to diagnose growth hormone (GH) deficiency and predict therapy response: first and second retesting during replacement therapy of patients defined as GH deficient, J. Clin. Endocrinol. Metab. 79 (1994) 1663–1669.
[9] M. Maghnie, C. Strigazzi, C. Tinelli, et al., Growth hormone (GH) deficiency (GHD) of childhood onset: reassessment of GH status and evaluation of the predictive criteria for permanent GHD in young adults, J. Clin. Endocrinol. Metab. 84 (1999) 1324–1328.
[10] R. Coutant, S. Rouleau, F. Despert, N. Magontier, D. Loisel, J.M. Limal, Growth and adult height in GH-treated children with non-acquired GH deficiency and idiopathic short stature: the influence of pituitary magnetic resonance imaging findings, J. Clin. Endocrinol. Metab. 86 (2001) 4649–4654.
[11] S.B.C. Loche, M. Maghnie, A. Faedda, C. Tzialla, M. Autelli, M.R. Casini, M. Cappa, Resuts of early reevaluation of growth hormone secretion in short children with apparent growth hormone deficiency, J. Pediatr. 140 (2002) 445–449.
[12] G.E. Butler, M. McKie, S.G. Ratcliffe, The cyclical nature of prepubertal growth, Ann. Hum. Biol. 17 (1990) 177–198.
[13] J. Van den Broeck, P. Hering, A. Van de Lely, A. Hokken-Koelega, Interpretative difficulties with growth hormone provocative retesting in childhood-onset growth hormone deficiency, Horm. Res. 51 (1999) 1–9.
[14] S. Longobardi, B. Merola, R. Pivonello, et al., Reevaluation of growth hormone (GH) secretion in 69 adults diagnosed as GH-deficient patients during childhood, J. Clin. Endocrinol. Metab. 81 (1996) 1244–1247.
[15] S. Wacharasindhu, A.M. Cotterill, C. Camacho-Hubner, G.M. Besser, M.O. Savage, Normal growth hormone secretion in growth hormone insufficient children retested after completion of linear growth, Clin. Endocrinol. (Oxf) 45 (1996) 553–556.
[16] H.G. Goodman, M.M. Grumbach, S.L. Kaplan, Growth and growth hormone. II. A comparison of isolated growth-hormone deficiency and multiple pituitary-hormone deficiencies in 35 patients with idiopathic hypopituitary dwarfism, N. Engl. J. Med. 278 (1968) 57–68.
[17] S.L. Kaplan, C.A. Abrams, J.J. Bell, F.A. Conte, M.M. Grumbach, Growth and growth hormone. I. Changes in serum level of growth hormone following hypoglycemia in 134 children with growth retardation, Pediatr. Res. 2 (1968) 43–63.
[18] V. Tillmann, J.M. Buckler, M.S. Kibirige, et al., Biochemical tests in the diagnosis of childhood growth hormone deficiency, J. Clin. Endocrinol. Metab. 82 (1997) 531–535.
[19] L. Adan, J.C. Souberbielle, R. Brauner, Diagnostic markers of permanent idiopathic growth hormone deficiency, J. Clin. Endocrinol. Metab. 78 (1994) 353–358.
[20] D.M. Hoffman, A.J. O'Sullivan, R.C. Baxter, K.K. Ho, Diagnosis of growth-hormone deficiency in adults, Lancet 343 (1994) 1064–1068.
[21] R. Youlton, S.L. Kaplan, M.M. Grumbach, Growth and growth hormone. IV. Limitations of the growth hormone response to insulin and arginine and of the immunoreactive insulin response to arginine in the assessment of growth hormone deficiency in children, Pediatrics 43 (1969) 989–1004.
[22] D.T. Wyatt, D. Mark, A. Slyper, Survey of growth hormone treatment practices by 251 pediatric endocrinologists, J. Clin. Endocrinol. Metab. 80 (1995) 3292–3297.
[23] G. Marin, H.M. Domene, K.M. Barnes, B.J. Blackwell, F.G. Cassorla, G.B. Cutler Jr., The effects of estrogen priming and puberty on the growth hormone response to standardized treadmill exercise and arginine-insulin in normal girls and boys, J. Clin. Endocrinol. Metab. 79 (1994) 537–541.
[24] E. Ghigo, J. Bellone, G. Aimaretti, et al., Reliability of provocative tests to assess growth hormone secretory status. Study in 472 normally growing children, J. Clin. Endocrinol. Metab. 81 (1996) 3323–3327.
[25] A. Rahim, A.A. Toogood, S.M. Shalet, The assessment of growth hormone status in normal young adult males using a variety of provocative agents, Clin. Endocrinol. (Oxf) 45 (1996) 557–562.
[26] H.J. Guyda, Growth hormone testing and the short child, Pediatr. Res. 48 (2000) 579–580.
[27] M.L. Granada, A. Sanmarti, A. Lucas, et al., Assay-dependent results of immunoassayable spontaneous 24-h growth hormone secretion in short children, Acta Paediatr. Scand. Suppl. 370 (1990) 63–70, discussion 71.
[28] N. Mauras, P. Walton, M. Nicar, S. Welch, A.D. Rogol, Growth hormone stimulation testing in both short and normal statured children: use of an immunofunctional assay, Pediatr. Res. 48 (2000) 614–618.
[29] A.M. Andersson, H. Orskov, M.B. Ranke, S. Shalet, N.E. Skakkebaek, Interpretation of growth hormone provocative tests: comparison of cut-off values in four European laboratories, Eur. J. Endocrinol. 132 (1995) 340–343.
[30] L. Cavallo, A. Acquafredda, S. Liuzzi, et al., Growth hormone release during insulin tolerance, clonidine, arginine and growth hormone releasing hormone tests in short normal children and adolescents, J. Endocrinol. Invest. 15 (1992) 131–135.
[31] M. Tzanela, H. Guyda, G. Van Vliet, G.S. Tannenbaum, Somatostatin pretreatment enhances growth hormone (GH) responsiveness to GH-releasing hormone: a potential new diagnostic approach to GH deficiency, J. Clin. Endocrinol. Metab. 81 (1996) 2487–2494.
[32] E.N. Gonc, N. Yordam, N. Kandemir, A. Alikasifoglu, Comparison of stimulated growth hormone levels in primed versus unprimed provocative tests. Effect of various testosterone doses on growth hormone levels, Horm. Res. 56 (2001) 32–37.
[33] A.S. Martinez, H.M. Domene, M.G. Ropelato, et al., Estrogen priming effect on growth hormone (GH) provocative test: a useful tool for the diagnosis of GH deficiency, J. Clin. Endocrinol. Metab. 85 (2000) 4168–4172.
[34] J.D. Miller, N.M. Wright, A. Esparza, et al., Spontaneous pulsatile growth hormone release in male and female premature infants, J. Clin. Endocrinol. Metab. 75 (1992) 1508–1513.
[35] M. Bozzola, K. Tettoni, F. Locatelli, et al., Postnatal variations of growth hormone bioactivity and of growth hormone-dependent factors, Arch. Pediatr. Adolesc. Med. 150 (1996) 1068–1071.
[36] A. Bhala, M.C. Harris, S. Zirin, L. Corcoran, P. Cohen, Insulin-like growth factor axis parameters in sick hospitalized neonates, J. Pediatr. Endocrinol. Metab. 11 (1998) 451–459.
[37] W.J. Smith, T.J. Nam, L.E. Underwood, W.H. Busby, A. Celnicker, D.R. Clemmons, Use of insulin-like growth factor-binding protein-2 (IGFBP-2), IGFBP-3, and IGF-I for assessing growth hormone status in short children, J. Clin. Endocrinol. Metab. 77 (1993) 1294–1299.
[38] G.M. Bright, J.R. Julius, J. Lima, S.L. Blethen, Growth hormone stimulation test results as predictors of recombinant human growth hormone treatment outcomes: preliminary analysis of the national cooperative growth study database, Pediatrics 104 (1999) 1028–1031.
[39] J. Van den Broeck, N. Arends, A. Hokken-Koelega, Growth response to recombinant human growth hormone (GH) in children with idiopathic growth retardation by level of maximum GH peak during GH stimulation tests, Horm. Res. 53 (2000) 267–273.
[40] F.R. Kaufman, K. Epport, R. Engilman, M. Halvorson, Neurocognitive functioning in children diagnosed with diabetes before age 10 years, J. Diabetes Complications 13 (1999) 31–38.
[41] E.A. Northam, P.J. Anderson, R. Jacobs, M. Hughes, G.L. Warne, G.A. Werther, Neuropsychological profiles of children

with type 1 diabetes 6 years after disease onset, Diabetes Care 24 (2001) 1541–1546.
[42] J.F. Rovet, R.M. Ehrlich, The effect of hypoglycemic seizures on cognitive function in children with diabetes: a 7-year prospective study, J. Pediatr. 134 (1999) 503–506.
[43] P.J. Galloway, E. McNeill, W.F. Paterson, M.D. Donaldson, Safety of the insulin tolerance test, Arch. Dis. Child 87 (2002) 354–356.
[44] A. Shah, R. Stanhope, D. Matthew, Hazards of pharmacological tests of growth hormone secretion in childhood, BMJ 304 (1992) 173–174.
[45] C. Huang, K. Banerjee, E. Sochett, K. Perlman, D. Wherrett, D. Daneman, Hypoglycemia associated with clonidine testing for growth hormone deficiency, J. Pediatr. 139 (2001) 323–324.
[46] E. Cacciari, P. Tassoni, G. Parisi, et al., Pitfalls in diagnosing impaired growth hormone (GH) secretion: retesting after replacement therapy of 63 patients defined as GH deficient, J. Clin. Endocrinol. Metab. 74 (1992) 1284–1289.
[47] Consensus guidelines for the diagnosis and treatment of growth hormone (GH) deficiency in childhood and adolescence: summary statement of the GH Research Society. GH Research Society. J. Clin. Endocrinol. Metab. 85 (2000) 3990–3993.
[48] S.D. Frasier, Editorial: the diagnosis and treatment of childhood and adolescent growth hormone deficiency – consensus or confusion?, J. Clin. Endocrinol. Metab. 85 (2000) 3988–3989.
[49] S. Cianfarani, S. Boemi, A. Spagnoli, et al., Is IGF binding protein-3 assessment helpful for the diagnosis of GH deficiency?, Clin. Endocrinol. (Oxf) 43 (1995) 43–47.
[50] S. Cianfarani, T. Tondinelli, G.L. Spadoni, G. Scire, S. Boemi, B. Boscherini, Height velocity and IGF-I assessment in the diagnosis of childhood onset GH insufficiency: do we still need a second GH stimulation test?, Clin. Endocrinol. (Oxf) 57 (2002) 161–167.
[51] H. Mitchell, M.T. Dattani, V. Nanduri, P.C. Hindmarsh, M.A. Preece, C.G. Brook, Failure of IGF-I and IGFBP-3 to diagnose growth hormone insufficiency, Arch. Dis. Child 80 (1999) 443–447.
[52] M.B. Ranke, R. Schweizer, M.W. Elmlinger, et al., Relevance of IGF-I, IGFBP-3, and IGFBP-2 measurements during GH treatment of GH-deficient and non-GH-deficient children and adolescents, Horm. Res. 55 (2001) 115–124.
[53] V. Tillmann, V.W. Tang, D.A. Price, D.G. Hughes, N.B. Wright, P.E. Clayton, Magnetic resonance imaging of the hypothalamic–pituitary axis in the diagnosis of growth hormone deficiency, J. Pediatr. Endocrinol. Metab. 13 (2000) 1577–1583.
[54] M. Bozzola, F. Mengarda, P. Sartirana, L. Tato, J.L. Chaussain, Long-term follow-up evaluation of magnetic resonance imaging in the prognosis of permanent GH deficiency, Eur. J. Endocrinol. 143 (2000) 493–496.
[55] D. Zenaty, C. Garel, C. Limoni, P. Czernichow, J. Leger, Presence of magnetic resonance imaging abnormalities of the hypothalamic–pituitary axis is a significant determinant of the first 3 years growth response to human growth hormone treatment in prepubertal children with nonacquired growth hormone deficiency, Clin. Endocrinol. (Oxf) 58 (2003) 647–652.
[56] M.G. Osorio, S. Marui, A.A. Jorge, et al., Pituitary magnetic resonance imaging and function in patients with growth hormone deficiency with and without mutations in GHRH-R, GH-1, or PROP-1 genes, J. Clin. Endocrinol. Metab. 87 (2002) 5076–5084.
[57] G. Binder, B.H. Nagel, M.B. Ranke, P.E. Mullis, Isolated GH deficiency (IGHD) type II: imaging of the pituitary gland by magnetic resonance reveals characteristic differences in comparison with severe IGHD of unknown origin, Eur. J. Endocrinol. 147 (2002) 755–760.
[58] T.A. Wilson, S.R. Rose, P. Cohen, A.D. Rogol, P. Backeljauw, R. Brown, D.S. Hardin, S.F. Kemp, M. Lawson, S. Radovick, S.M. Rosenthal, L. Silverman, P. Speiser, Update of guidelines for the use of growth hormone in children: The Lawson Wilkins pediatric endocrinology society drug and therapeutics committee, J. Pediatr. 143 (2003) 415–421.