Distinguishing Constitutional Delay of Growth and Puberty from Isolated Hypogonadotropic Hypogonadism: Critical Appraisal of Available Diagnostic Tests

Jennifer Harrington and Mark R. Palmert
Division of Endocrinology, The Hospital for Sick Children and Department of Pediatrics, The University of Toronto, Toronto, Canada M5G1X8

Context: Determining the etiology of delayed puberty during initial evaluation can be challenging. Specifically, clinicians often cannot distinguish constitutional delay of growth and puberty (CDGP) from isolated hypogonadotropic hypogonadism (IHH), with definitive diagnosis of IHH awaiting lack of spontaneous puberty by age 18 yr. However, the ability to make a timely, correct diagnosis has important clinical implications.

Objective: The aim was to describe and evaluate the literature regarding the ability of diagnostic tests to distinguish CDGP from IHH.

Evidence Acquisition: A PubMed search was performed using key words “puberty, delayed” and “hypogonadotropic hypogonadism,” and citations within retrieved articles were reviewed to identify studies that assessed the utility of basal and stimulation tests in the diagnosis of delayed puberty. Emphasis was given to a test’s ability to distinguish prepubertal adolescents with CDGP from those with IHH.

Evidence Synthesis: Basal gonadotropin and GnRH stimulation tests have limited diagnostic specificity, with overlap in gonadotropin levels between adolescents with CDGP and IHH. Stimulation tests using more potent GnRH agonists and/or human chorionic gonadotropin may have better discriminatory value, but small study size, lack of replication of diagnostic thresholds, and prolonged protocols limit clinical application. A single inhibin B level in two recent studies demonstrated good differentiation between groups.

Conclusion: Distinguishing IHH from CDGP is an important clinical issue. Basal inhibin B may offer a simple, discriminatory test if results from recent studies are replicated. However, current literature does not allow for recommendation of any diagnostic test for routine clinical use, making this an important area for future investigation. (J Clin Endocrinol Metab 97: 3056–3067, 2012)

Delayed puberty is defined as the absence of signs of sexual maturation by an age more than 2–2.5 SD values above the mean of the population (traditionally breast development by 13 yr in girls and testicular development by 14 yr in boys) (1, 2). Delayed puberty can stem from conditions that cause hypergonadotropic hypogonadism (such as Turner syndrome or chemotherapy-induced gonadal toxicity), permanent hypogonadotropic hypogonadism (HH) [such as multiple pituitary hormone deficiency (MPHD), Kallmann syndrome, or high-dose central nervous system irradiation], or the transient HH that is seen in systemic conditions (such as anorexia or inflammatory bowel syndrome) (3). A careful history, physical examination, and diagnostic evaluation can often identify these underlying disorders (4). However, in a significant proportion of cases (at least 65% of males and 45% of females) (5), the etiology remains unidentifiable.
30% of females), delayed puberty represents the extreme end of the distribution of normal timing of puberty, rather than overt pathology (3). This normal variation of growth and development, termed constitutional delay of growth and puberty (CDGP), is the single most common cause of delayed puberty in both genders.

CDGP is a diagnosis of exclusion. It is often clinically challenging to differentiate adolescents with CDGP from those with a form of HH termed isolated HH (IHH), which usually is a permanent condition (5, 6). Distinguishing between these conditions is especially difficult during initial evaluations because adolescents with both etiologies are often prepubertal on examination and have low levels of gonadotropins (LH and FSH). LH and FSH levels are low in CDGP because the hypothalamic-pituitary-gonadal (HPG) axis has not yet matured to secrete pubertal levels of GnRH; levels are low in IHH because of a lack of GnRH secretion or action (7, 8).

There are clinical features that can potentially distinguish CDGP from IHH, although these are often not diagnostic (4). A family history of delayed puberty is strongly suggestive of CDGP (seen in 50–75%) (3, 9, 10), although individuals with CDGP are also seen among pedigrees with IHH (11, 12). Adolescents with CDGP may have delayed adrenarche and pubarche along with delayed gonadal development, whereas individuals with IHH are more likely to have delayed gonadal development alone (3), but this distinction is often blurred. Twenty to 40% of patients with IHH have evidence of initial but then stalled pubertal development (11, 12), and 10% may have sustained reversal of IHH in adulthood (5, 6). Small testes on examination (1–2 ml in volume), a history of undescended testes, and/or small phallus can point toward a diagnosis of IHH. However, the prevalence of cryptorchidism in IHH varies between 5 and 40%, and in those patients with IHH with evidence of partial pubertal development, it approximates that of the general population (3–5% at birth) (11, 12). Anosmia or hyposmia, a feature in Kallmann syndrome, occurs in only 30–50% of patients with apparent IHH (12, 13).

Thus, in many cases, routine initial clinical evaluation cannot distinguish CDGP from IHH with certainty. The presence of endogenous, progressive pubertal development by age 18 yr is the “gold standard” for differentiating CDGP from IHH. The inability to make an accurate diagnosis at initial presentation presents difficulties in providing appropriate counseling around prognosis, may generate anxiety among adolescents and families, and can affect management decisions. Endocrinologists have long sought a diagnostic test for CDGP, and different physiological and stimulation tests have been proposed as early discriminators (Fig. 1). In this review we describe and evaluate the literature regarding diagnostic tests proposed to distinguish CDGP from HH, in particular IHH.

**Diagnostic Tests**

We identified a total of 19 studies published in English over the past 30 yr whose primary objective was to assess the sensitivity of a diagnostic test in the differentiation of HH from CDGP in adolescents. Where available, emphasis has been given to data specifically comparing adolescents with CDGP and IHH. When possible, IHH is used instead of the more general HH to distinguish this patient population from adolescents with MPHD, who have been included in many studies. It is important to note that the majority only involved the investigation of boys with delayed puberty, and thus conclusions drawn from the data...
may have limitations in their application to the investigation of delayed puberty in girls.

**Basal gonadotropin levels (Table 1)**

In childhood, there are low but detectable levels of GnRH-stimulated LH and FSH pulses, occurring predominantly at night, which progressively increase in number and amplitude with the onset of puberty (14). It has been hypothesized that gonadotropin levels in individuals with IHH may be lower than those seen in CDGP, even before the onset of puberty. Several studies have examined whether there are diagnostic basal gonadotropin values that could discriminate between CDGP and IHH.

Frequent sampling for the presence or absence of LH nocturnal pulses has been proposed as a distinguishing characteristic, with the lack of nocturnal LH pulses in adolescence being initially described as highly specific for HH (15). However, using ultrasensitive assays with lower limits of detection, in many adolescents and young adults with IHH, low-amplitude pulses of gonadotropins are detectable. The LH and FSH pulses are at levels indistinguishable from those in prepubertal controls, although the pulses do not have the same association with sleep onset (16). Whether there are differences in the pattern of nocturnal gonadotropin secretion between adolescents with CDGP and IHH using ultra-sensitive assays has not yet been determined.

Given that overnight, frequent sampling is not a practical routine diagnostic test, studies have assessed the diagnostic utility of a single basal gonadotropin level. As a group, patients with IHH have significantly lower basal gonadotropins than age- and Tanner stage-matched children with CDGP, but there is substantial overlap between groups (17, 18).

A recent study reported that a basal FSH level below 1.2 IU/liter in boys presenting with delayed puberty confirmed the diagnosis of HH with 100% specificity, precluding the need for further testing (19). This conclusion was, however, based upon the results of only seven patients with CDGP, with FSH levels ranging from 1.2–3.4 IU/liter. Other groups have reported lower FSH levels in subjects with CDGP and have not been able to identify a diagnostic FSH threshold to distinguish CDGP from HH (20, 21). FSH levels in large cohorts also have significant overlap between prepubertal and early-pubertal children (22). Thus, the diagnostic utility of a basal FSH level remains unclear.

LH levels are a more sensitive indicator for the onset of central puberty than FSH (23). Sequera et al. (24) demonstrated that a basal LH of more than 0.65 IU/liter excluded a diagnosis of complete HH, defined as testes volume of less than 3 ml on follow-up at 20 yr of age. Subjects with partial HH (testes volume of 6–12 ml at follow-up) could not be identified using any basal LH threshold. These results are again based upon small sample sizes; however, in this instance, similar data have been reported in other studies, although with variability in the exact threshold LH value due to assay differences (19, 20). Validation of threshold levels with reliable, cross-checked assays is crucial for the test’s clinical use. Even if validated, there is limited utility in having a diagnostic test that only excludes complete, but not partial

### TABLE 1. Studies that have used basal gonadotropin levels to diagnose HH

<table>
<thead>
<tr>
<th>Study</th>
<th>First author, year (Ref.)</th>
<th>Test protocol (assay method for gonadotropins)</th>
<th>HH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wu, 1991 (16)</td>
<td>12-h overnight sampling for LH and FSH pulses (IFMA)</td>
<td>No. of subjects: 8 males; Age (yr): 24 (11–36); Testes volume or length: 1.6 (±1.4) ml</td>
</tr>
<tr>
<td>2</td>
<td>Odink, 1998 (15)</td>
<td>24-h sampling for LH pulses (IRMA)</td>
<td>No. of subjects: 14 males, 11 females (1/3 with MPHD); Age (yr): 16.4 (9.9–20.6); Testes volume or length: 2 ml</td>
</tr>
<tr>
<td>3</td>
<td>Sequera, 2002 (24)</td>
<td>Basal LH and FSH (IFMA)</td>
<td>No. of subjects: 12 males (5 partial HH, 7 complete HH); Age (yr): 16.8 (14.8–18.4); Testes volume or length: 3.9 (1–10) ml</td>
</tr>
<tr>
<td>4</td>
<td>Grinspon, 2010 (19)</td>
<td>Basal LH and FSH (IFMA)</td>
<td>No. of subjects: 25 males (40% MPHD); Age (yr): 16.4 (±3.1); Testes volume or length: 1.8 (±0.6) ml</td>
</tr>
</tbody>
</table>

Data are expressed as mean (±SD or range). General lower limits of detection of LH assays: RIA, 0.5–2 IU/liter; IRMA, 0.1–0.5 IU/liter; IFMA and ICMA, <0.1 IU/liter. PPV, Positive predictive value; NPV, negative predictive value; IFMA, immunofluorometric assay; IRMA, immunoradiometric assay; ICMA, immunochemiluminescence assay.
IHH. LH levels of more than 0.65 IU/liter are high enough that individuals who meet these criteria may already be in puberty clinically, and the biochemical data may not add substantially to the diagnosis.

Thus, whereas on average adolescents with IHH have lower basal gonadotropin levels than those with CDGP, basal gonadotropins appear to have limited discriminatory ability for an individual patient.

**GnRH and GnRH agonist (GnRHa) stimulation tests (Table 2)**

Given the diagnostic limitations of basal gonadotropins, there have been many attempts to use stimulation tests to distinguish between CDGP and IHH. For example, an iv bolus of GnRH leads to a dose-dependent increase in gonadotropins (25), both in prepubertal and pubertal adolescents (26). LH responsiveness to GnRH stimulation increases significantly after a bone age of 10–11 yr in healthy boys reflecting maturation of the HPG axis (27). An underlying premise is that this response to GnRH (or GnRHa) will be more robust in individuals with CDGP (who have been previously exposed to endogenous GnRH) than in individuals with IHH (who have not been exposed to GnRH or who have inactive GnRH receptors).

Unfortunately, significant variability in response has led to an inability to distinguish accurately between the two groups. A proportion of adolescents with CDGP, even with bone ages greater than 11 yr, demonstrate low peak LH levels on GnRH stimulation testing, reflecting the spectrum of maturation of the HPG axis (26, 28). This variability is also observed in the general population with overlap in stimulated LH levels between Tanner stages 1 and 2 and stages 2 and 3 (29, 30). As a group, adolescents with HH have lower stimulated LH levels compared with subjects with CDGP, but several studies have demonstrated that up to 30% have LH responses indistinguishable from those with CDGP (31, 32). Overlap between CDGP and HH groups is seen whether GnRH is given as a bolus or as an infusion over 120 min (19, 24).

Variations on the GnRH test have been studied in trials to improve its diagnostic utility, including the administration of repetitive small doses of iv GnRH for 36 h before a traditional GnRH stimulation test. Pulsatile GnRH leads to a decreased stimulated LH response in subjects with CDGP and IHH (33). The decrease in stimulated LH is greater in the IHH group, possibly reflecting more pronounced depletion of fast releasable LH stores (34). A small study of 17 subjects reported clear differentiation between adolescents with CDGP and IHH using this method (33). Because the goal is to distinguish prepubertal patients with CDGP from prepubertal patients with IHH, problematically eight of the nine in the CDGP group were clinically in puberty at the time of testing (testes volume >4 ml). A second study also showed that this method was better than the traditional GnRH test alone but did not demonstrate 100% diagnostic accuracy (34). If GnRH pulses are administered for longer than 36 h, gonadotropin secretion increases in both CDGP and IHH, with no difference in post-GnRH stimulation testing between groups (35).

**TABLE 1. Continued**

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of subjects</th>
<th>Age (yr)</th>
<th>Testes volume or length</th>
<th>Study results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16 males, 6 females, prepubertal</td>
<td>6.2 (4.4–8.1)</td>
<td>2 ml</td>
<td>No significant difference in frequency or amplitude of LH or FSH pulses between groups. IHH group had an absence of sleep entrainment of LH pulses.</td>
</tr>
<tr>
<td>2</td>
<td>6 males and 2 females, CDGP</td>
<td>15.8 (13.5–18.9)</td>
<td>3.9 (2–8) ml</td>
<td>Lack of LH pulses had a 74% PPV and 100% NPV for identification of HH. A basal FSH level ≤1.11 IU/liter in males and ≤2.86 IU/liter in females had a 97–100% sensitivity, 23–28% specificity for lack of LH pulses.</td>
</tr>
<tr>
<td>3</td>
<td>8 males, CDGP</td>
<td>15.4 (14–17.3)</td>
<td>4.4 (2–10) ml</td>
<td>A single basal LH &gt;0.65 IU/liter excluded complete HH. Basal gonadotropins unable to distinguish between CDGP and partial HH (complete HH defined in this study as testes volume &lt;4 ml and partial HH as testes volume 6–12 ml after 5 yr of clinical follow-up).</td>
</tr>
<tr>
<td>4</td>
<td>7 males, CDGP</td>
<td>13.9 (±1.8)</td>
<td>2.9 (±1.0) ml</td>
<td>A single basal FSH level of &lt;1.2 IU/liter had a 100% PPV and 54% NPV to predict HH.</td>
</tr>
</tbody>
</table>
the administration of GnRH pulses is complicated and invasive, precluding it from being used routinely.

Another more applicable technique has been the use of GnRHa stimulation tests. Different GnRHa are characterized by alterations in the amino acid in position 6 (36), resulting in increased potency and half-life due to greater affinity for the GnRH receptor and resistance to enzyme degradation (37). These effects have been postulated to allow better distinction between CDGP and IHH because a greater stimulus would result in “awakening” of primed normal gonadotropic cells in patients with CDGP (38).

**Nafarelin**

In a pilot study of three patients with CDGP and eight patients with HH, nafarelin completely discriminated between diagnostic groups (38). A limitation, in addition to the small cohort size, was that the patients with HH were distinguishable on clinical grounds alone, given the presence of other pituitary hormone deficiencies, anosmia, or micropenis. Subsequent studies have demonstrated reasonable discriminatory ability using the nafarelin stimulation test but with overlap in LH increment between groups (39, 40). Of note when comparing studies, the LH increment in the patients with HH was significantly different despite the same test protocol being used. There have not been any recently published studies of this stimulation method using ultra-sensitive LH assays. From the currently published studies, a consistent diagnostic threshold to distinguish between CDGP and IHH cannot be determined, and so further research is needed before definite conclusions regarding general clinical utility can be drawn.

**Triptorelin**

Triptorelin stimulation testing has been shown in two studies to differentiate completely between CDGP and HH, with all HH patients having a peak LH level of less than 9 IU/liter 4 h after iv injection (41, 42). Zamboni et al. (41), who included a second prepubertal control group, found that this differentiation only occurred if the pa-

### TABLE 2. Studies that have used GnRH and GnRHa stimulation tests to diagnose HH

<table>
<thead>
<tr>
<th>Study</th>
<th>First author, year (Ref.)</th>
<th>Test protocol (assay method for gonadotropins)</th>
<th>HH No. of subjects</th>
<th>Age (yr)</th>
<th>Testes volume or length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dunkel, 1985 (32)</td>
<td>GnRH iv bolus, 3.5 µg/kg (RIA)</td>
<td>21 males (12 HH, 9 MPHD)</td>
<td>17.4 (12.5–23.4)</td>
<td>3.9 (0.8–9.6) ml</td>
</tr>
<tr>
<td>2</td>
<td>Sequera, 2002 (24)</td>
<td>GnRH iv 100-µg infusion over 120 min (IFMA)</td>
<td>12 males (5 partial HH, 7 complete HH)</td>
<td>16.8 (14.8–18.4)</td>
<td>3.9 (1–10) ml</td>
</tr>
<tr>
<td>3</td>
<td>Grinspon, 2010 (19)</td>
<td>GnRH iv 100-µg infusion over 120 min (IFMA)</td>
<td>25 males (40% MPHD)</td>
<td>16.4 (±3.1)</td>
<td>1.8 (±0.6) ml</td>
</tr>
<tr>
<td>4</td>
<td>Partsch, 1985 (33)</td>
<td>GnRH pump (36 h of 5 µg every 90 min)</td>
<td>8 males</td>
<td>20.9 (15.5–41)</td>
<td>3 (1–4) ml</td>
</tr>
<tr>
<td>5</td>
<td>Smals, 1994 (34)</td>
<td>GnRH iv stimulation test 60 µg/m² after pump (RIA)</td>
<td>16 males</td>
<td>19.3 (15–24)</td>
<td>2.4 (1–4) cm</td>
</tr>
<tr>
<td>6</td>
<td>Ehrmann, 1989 (38)</td>
<td>Nafarelin 1 µg/kg sc (RIA)</td>
<td>8 males (2 MPHD)</td>
<td>18.1 (14.3–24)</td>
<td>1.6 (1–3) cm</td>
</tr>
<tr>
<td>7</td>
<td>Ghai, 1995 (40)</td>
<td>Nafarelin 1 µg/kg sc (RIA)</td>
<td>10 males (5 MPHD)</td>
<td>15.1 (13.3–16.8)</td>
<td>1 (1–2) cm</td>
</tr>
<tr>
<td>8</td>
<td>Kletter, 1996 (39)</td>
<td>Nafarelin 1 µg/kg up to 100 µg sc (RIA)</td>
<td>4 males, 1 female</td>
<td>19.3 (17.5–24)</td>
<td>&lt;3 ml</td>
</tr>
<tr>
<td>9</td>
<td>Zamboni, 1995 (41)</td>
<td>Triptorelin 0.1 mg/m² sc (IFMA)</td>
<td>10 males (6 MPHD)</td>
<td>16.8 (15.2–18.9)</td>
<td>2.8 (2–4) ml</td>
</tr>
<tr>
<td>10</td>
<td>Kauschansky, 2002 (42)</td>
<td>Triptorelin 0.1 mg/m² sc (ICMA)</td>
<td>19 males</td>
<td>16.1 (14–18)</td>
<td>1–3 ml</td>
</tr>
<tr>
<td>11</td>
<td>Degros, 2003 (43)</td>
<td>Triptorelin 0.1 mg sc (MEIA)</td>
<td>13 males</td>
<td>19.9 (±3.3)</td>
<td>2.7 (±1.6) ml</td>
</tr>
<tr>
<td>12</td>
<td>Wilson, 2006 (18)</td>
<td>Buserelin 100 µg sc (ICMA)</td>
<td>8 males (1 MPHD, 1 GHD)</td>
<td>13.2 (10.3–14.3)</td>
<td>1 (0–2) ml</td>
</tr>
<tr>
<td>13</td>
<td>Street, 2002 (45)</td>
<td>Leuprolide 500 µg sc (RIA)</td>
<td>11 males</td>
<td>16.5 (13–29.3)</td>
<td>2 (1–3) ml</td>
</tr>
</tbody>
</table>

Data are expressed as mean (so or range). PPV, Positive predictive value; NPV, negative predictive value; IFMA, immunofluorometric assay; IRMA, immunoradiometric assay; ICMA, immunochemiluminescence assay; MEIA, microparticle enzyme immunoassay; GHD, GH deficiency.
tient’s bone age was greater than 12 yr. In contrast, a third study found significant overlap in LH levels after triptorelin stimulation between groups (43). The difference in results between studies may be related to differences in assays or in the CDGP groups used [entering puberty within 12 months of testing (41, 42) compared with later onset of pubertal development (43)].

Buserelin

Using another GnRHa, buserelin, Wilson et al. (18) demonstrated that a peak LH level below 5 IU/liter after stimulation had an 89% positive and 100% negative predictive value in identifying the patients with HH. A significant limitation was that the patient groups were not age-matched. The HH subjects were significantly younger than the CDGP subjects, and several of the HH subjects did not meet criteria for having delayed puberty. Although promising, these results have not been replicated in the literature.

Leuprolide

After an initial trial that proposed that leuprolide stimulation testing may be a better predictor of pubertal progression than GnRH testing (44), there have been two studies assessing this specifically in adolescents with CDGP and IHH. Street et al. (45) demonstrated a 100% sensitivity and specificity to distinguish between diagnostic groups, with all patients with IHH having a stimulated LH of 2.8 IU/liter or less. Unfortunately, Lanes et al. (18) did not demonstrate the same diagnostic utility of leuprolide stimulation testing; the observed LH values had significant overlap between diagnostic groups (46).

Summary

GnRHa appear to offer better discriminatory value than GnRH stimulation testing, with more robust results being replicated in the potent agonists (nafarelin and triptorelin). With the exception of leuprolide, the tests take a longer period of time to perform. Although promising, the

---

**TABLE 2. Continued**

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of subjects</th>
<th>Age (yr)</th>
<th>Testes volume or length</th>
<th>Study results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GnRH stimulation test 52 males, CDGP</td>
<td>16.1 (13.3–19.7)</td>
<td>1–10 ml</td>
<td>Reference stimulated LH values were calculated using the 95% confidence interval for the CDGP group at each genital Tanner stage. 13 of 21 patients with HH had a low peak LH level; 100% PPV, 87% NPV for GnRH test to identify patient with HH.</td>
</tr>
<tr>
<td>2</td>
<td>8 males, CDGP</td>
<td>15.4 (14–17.3)</td>
<td>4.4 (2–10) ml</td>
<td>Overlap in peak LH between HH and CDGP. Peak LH: complete HH, 1.1 to 6; partial HH, 3.04 to 30.8; CDGP, 3.2 to 33.4 IU/liter. (Complete HH defined in this study as testes volume &lt;4 ml and partial HH as testes volume 6–12 ml after 5 yr of clinical follow-up.)</td>
</tr>
<tr>
<td>3</td>
<td>7 males, CDGP</td>
<td>13.9 (±1.8)</td>
<td>2.9 (±1.0) ml</td>
<td>Peak LH &lt;5.8 and peak FSH of &lt;4.6 IU/liter gave a 100% PPV and 64% NPV for HH.</td>
</tr>
<tr>
<td>4</td>
<td>GnRH stimulation tests after repetitive doses of GnRH 9 males, CDGP</td>
<td>16.3 (14.5–20)</td>
<td>7 (2–15) ml</td>
<td>No overlap in LH increment after GnRH testing between the groups. LH increment in HH group, 0.8 to 2.4; CDGP group, 4.1 to 15.6 IU/liter.</td>
</tr>
<tr>
<td>5</td>
<td>17 males, CDGP</td>
<td>16.5 (14–21.5)</td>
<td>2.5 (1.5–3.5) cm</td>
<td>A LH increment after GnRH stimulation of ≤3 had an 89% PPV, 100% NPV for HH.</td>
</tr>
<tr>
<td>6</td>
<td>GnRHa stimulation test 3 males, CDGP</td>
<td>16.2 (14.8–17.6)</td>
<td>2.2 (2–2.5) cm</td>
<td>No overlap in peak LH response between groups. Peak LH, HH, 5.5 (≥0.8) IU/liter; CDGP, 77.2 (≤8.6) IU/liter.</td>
</tr>
<tr>
<td>7</td>
<td>11 prepubertal, CDGP</td>
<td>14.9 (13.8–17.6)</td>
<td>1.7 (1–2) cm</td>
<td>Peak LH &lt;7.2 IU/liter had a 100% PPV, 95% NPV for HH.</td>
</tr>
<tr>
<td>8</td>
<td>11 early puberty, CDGP</td>
<td>15.4 (13.9–17.1)</td>
<td>2.8 (2.4–3.6) cm</td>
<td>LH increment after stimulation: HH, 0–6.0 IU/liter; CDGP, 4.8–49.2 IU/liter.</td>
</tr>
<tr>
<td>9</td>
<td>6 males, CDGP</td>
<td>15.3 (14.1–15.8)</td>
<td>&lt;3 ml</td>
<td>Overlap in LH increment after nafarelin between groups. PPV, NPV not reported.</td>
</tr>
<tr>
<td>10</td>
<td>18 males, CDGP</td>
<td>15.8 (15–17)</td>
<td>3.1 (2–4) ml</td>
<td>No overlap in peak LH between CDGP and HH groups, but complete overlap between prepubertal controls and HH.</td>
</tr>
<tr>
<td>11</td>
<td>16 prepubertal males</td>
<td>9.3 (6.9–11)</td>
<td>2.2 (2–3) ml</td>
<td>Peak LH results: HH, 0.1–8.6 IU/liter; CDGP, 13.5–38.1 IU/liter; prepubertal, 0.1–8.8 IU/liter.</td>
</tr>
<tr>
<td>12</td>
<td>13 males, CDGP</td>
<td>15.4 (14–21)</td>
<td>0.8–3 ml</td>
<td>No overlap in peak LH between CDGP and HH groups. Peak LH results: HH, 0.7–6.9 IU/liter; CDGP, 10.8–32.6 IU/liter.</td>
</tr>
<tr>
<td>13</td>
<td>19 males, CDGP</td>
<td>15.3 (±1.0)</td>
<td>4.8 (±1.8) ml</td>
<td>Peak LH level, &lt;14 IU/liter had a 72% PPV and 100% NPV to identify HH. Peak LH results: HH, 3.4 ± 4.1 IU/liter; CDGP, 18.4 ± 9.4 IU/liter.</td>
</tr>
<tr>
<td>14</td>
<td>23 males, CDGP (1 MPHID, 3 GHID)</td>
<td>14.6 (12.8–17.2)</td>
<td>2 (2–3) ml</td>
<td>All patients with HH had a peak LH &lt;5 IU/liter compared to 1 of 24 with CDGP.</td>
</tr>
<tr>
<td>15</td>
<td>7 males, CDGP</td>
<td>14.3 (13.5–15.3)</td>
<td>2.6 (2–3) ml</td>
<td>No overlap in peak LH levels 120–180 min after leuprolide between HH and CDGP groups, but overlap between prepubertal controls and HH. Peak LH results: HH, 0.7–2.8 IU/liter; CDGP, 6.1–15 IU/liter.</td>
</tr>
<tr>
<td>16</td>
<td>6 prepubertal males</td>
<td>9.5 (7.5–12.5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
low subject numbers in the studies and lack of replication of consistent diagnostic thresholds limits the ability to assess the ideal GnRHa to use. In addition, the paucity of any studies performed in females restricts conclusions to male patients.

**Human chorionic gonadotropin (hCG) test (Table 3)**

The hCG stimulation test has been used for many years, primarily to assess for the presence of functioning testicular tissue and to investigate defects of testosterone biosynthesis and action (47). The test is based on the ability of hCG to increase androgen production in Leydig cells via stimulation of the LH receptor (48). The normal Leydig response to hCG is thought to be dependent on previous exposure or “priming” by gonadotropins (49, 50). The deficiency of gonadotropins seen in patients with IHH would, theoretically, lead to a blunted testosterone response after hCG stimulation, analogous to the rationale behind the GnRH/GnRHa tests. Thus, the hCG test has been proposed to differentiate patients with CDGP from those with IHH.

Multiple protocols with different hCG doses, number of injections and sequence of blood draws for testosterone have been published (51–54), making the comparison of studies problematic. Several studies have reported predictive values of 82–86% of the hCG test to distinguish patients with HH from those with CDGP (20, 32, 43, 55).

Extending the hCG test to 19 d, Segal et al. (20) improved the positive predictive value to 92%.

Kauschansky et al. (42) reported even better results—100% sensitivity and specificity for the hCG test—using three injections of 1500 IU of hCG. The patients with CDGP in the study had significantly elevated basal LH levels compared with the IHH group (CDGP, LH, 0.7–2.0 IU/liter; vs. IHH, LH, 0.1–0.7 IU/liter using immunochemiluminescence assay), suggesting that some of the CDGP subjects were in early puberty at the time of testing. The very positive results in this study still need to be validated.

Combining the GnRH/GnRHa and hCG stimulation tests to improve diagnostic sensitivity has been investigated (32). One such study reported a 100% sensitivity and specificity for the hCG test—using three injections of 1500 IU of hCG. The patients with CDGP in the study had significantly elevated basal LH levels compared with the IHH group (CDGP, LH, 0.7–2.0 IU/liter; vs. IHH, LH, 0.1–0.7 IU/liter using immunochemiluminescence assay), suggesting that some of the CDGP subjects were in early puberty at the time of testing. The very positive results in this study still need to be validated.

**Inhibin B (Table 3)**

In addition to stimulation tests, researchers have investigated whether baseline measurement of gonadal prod-
Two studies have assessed the diagnostic utility of inhibin B in this clinical setting. Coutant et al. (17) demonstrated that a single inhibin B level of 35 pg/ml or less had a 93% positive predictive value to identify patients with IHH from those with CDGP. The predictive value increased to 100% when only assessing the patients with IHH who had testes volumes of less than 3 ml. However, the sensitivity and specificity of inhibin B was lower when comparing patients with HH as part of MPHD to the CDGP group. The second study by Adan et al. (65), using a significantly higher cutoff inhibin B level of less than 100 pg/ml, showed only a 73% positive predictive value to identify the boys with IHH. These differing results were found despite both studies using the same inhibin B assay and assessing patients with IHH at a similar age and pubertal stage.

A single inhibin B level may, with further verification, prove to be a simple first-line test in the diagnosis of delayed puberty in boys, with very low levels indicating a high likelihood of IHH. As with many of the other discriminatory tests, further comparative studies with clarification of diagnostic thresholds are needed to see whether the sensitivity and specificity of a single inhibin B test can be applied to routine clinical practice. However, it is encouraging that none of the boys with CDGP in either study had an inhibin B level of less than 35 pg/ml. A limitation in its use may be in its sensitivity to identify those boys with IHH who have a partial deficiency in gonadotropins, and therefore a partial deficiency in inhibin B.

The normal pattern of inhibin B secretion in females is similar to what is seen in males, but with lower prepubertal levels and a less significant rise at the time of puberty (66). Although there are no studies in female adolescents looking at the diagnostic utility of inhibin B, there are varying results in adults, with both low (67) and normal inhibin B levels (68) reported in adult females with HH compared with healthy controls. Further investigation is needed to determine whether inhibin B levels can also be used as a diagnostic test in females with delayed puberty.

**Genetic testing**

Despite remarkable progress made in identifying genes that cause IHH and that encode critical components of the HPG axis (7, 69, 70), mutations in known genes explain...
only 30–40% of IHH and Kallmann syndrome cases (7). CDGP has a strong genetic basis, with a family history of delayed puberty in 50–75% (3, 9, 10), but its genetic basis remains even more obscure. Because IHH is reversible in approximately 10% of adults who have been treated with testosterone (5, 6) and because some pedigrees with IHH include individuals with delayed but spontaneous puberty, it has been postulated that mutations in IHH-related genes might underlie CDGP. However, screening subjects with CDGP has thus far identified mutations only in rare, individual cases (71, 72). Analogously, 32 loci have recently been identified that modulate the age of menarche in the general population (73–77). Although not all of these have been studied in CDGP, sequencing has not identified any mutations in the most robustly associated gene, LIN28B (78).

There may be other forms of inheritance that underlie CDGP, including rarer variants with large phenotypic effects; combinations of variants within a single gene or multiple genes (oligogenicity); structural variation, such as copy number variants; and epigenetics. Some of these mechanisms have been identified as causes of hypothalamic amenorrhea (79). However with the current limitations in our understanding of the genetic basis underpinning both IHH and CDGP, routine genetic testing, particularly for diagnostic differentiation, is not warranted.

**Potential Implications in Distinguishing CDGP from IHH**

The natural history of adolescents presenting with delayed puberty will reveal the diagnosis, with complete and spontaneous pubertal progression eventually seen in CDGP, compared with incomplete or absent progression in IHH. There are, however, potential advantages in making an earlier diagnosis. Diagnostic uncertainty is associated with increased psychological stress for both adolescents and parents (80). The ability to make a diagnosis near the time of initial presentation and provide appropriate counseling may lead to reduced anxiety and eliminate the need for subsequent testing.

In addition, pubertal induction using gonadotropins and hCG has been proposed to potentially offer advantages in boys with IHH, compared with the use of testosterone alone (81, 82). Testosterone replacement only causes virilization, whereas treatment with hCG and recombinant FSH has been shown to induce testes growth, increase inhibin B levels (suggesting proliferation of immature Sertoli cells), and stimulate spermatogenesis in adolescent males with IHH (82, 83). Whether pubertal induction with gonadotropins in boys with IHH would be a positive predictor for future fertility still needs to be systematically studied, especially given the previously mentioned concerns about hCG use in infants with cryptorchidism (56–58), as well as the increased invasiveness and cost associated with this form of pubertal induction. However, if clinically significant differences in fertility are identified, this increases the importance of making an early diagnosis of IHH.

**Conclusions**

Over the past 30 yr, different basal and stimulation tests have been proposed to discriminate between adolescents with CDGP and IHH. Basal gonadotropins, genetic testing, GnRH and hCG stimulation tests all have limitations in diagnostic specificity and sensitivity to distinguish between groups. The more potent GnRHa appear to offer better discrimination, but validation of diagnostic thresholds and larger studies are needed. The complexity and invasiveness of the 36-h GnRH stimulation protocol precludes it from everyday clinical use. Conversely, recent studies indicate that inhibin B may provide a simple first-line test. If these initial reports are validated, this test may be able to identify a subset of patients with delayed puberty who have a high likelihood of having IHH.

It is important to appreciate some of the limitations inherent to the data that have been published. Variations in the type and reliability of assays used makes comparison of studies problematic. Although girls presenting with delayed puberty are more likely than males to have IHH, there are very few studies assessing the utility of diagnostic tests in females. Given the relative rarity of IHH, patients with HH as part of MPHID have often been studied. Although the inclusion of patients with MPHID allows demonstration of proof of principle for a diagnostic test, the need for discriminatory diagnostic tests in this patient population is lower. Similarly, the inclusion of control subjects who on clinical grounds alone would be readily identifiable as having CDGP (e.g. testes >6 ml with recent accelerated growth velocity) may give biased results. The clinical challenge for physicians is to diagnose the otherwise well prepubertal 14- to 15-yr-old adolescent.

Thus, the evidence-based literature regarding the available tests is insufficient to recommend any of them for routine clinical use. Further validation of previously published diagnostic thresholds as well as prospective studies focusing on prepubertal 14- to 15-yr-old adolescents are still needed.
Acknowledgments

We gratefully acknowledge Dr. Leo Dunkel for reviewing the manuscript and providing helpful comments.

Address all correspondence and requests for reprints to: Jennifer Harrington, Division of Endocrinology, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G1X8. E-mail: jennifer.harrington@sickkids.ca.

Disclosure Summary: The authors have nothing to disclose.

References

11. Piteloud N, Hayes FJ, Boepple PA, DeCruz S, Seminara SB, MacLaughlin DT, Crowley Jr WF 2002 The role of prior pubertal development, biochemical markers of testicular maturation, and genetics in elucidating the phenotypic heterogeneity of idiopathic hypogonadotropic hypogonadism. J Clin Endocrinol Metab 87: 152–160
29. Resende EA, Lara BH, Reis JD, Ferreira BP, Pereira GA, Borges MF


62. Nachtigall LB, Boepple PA, Seminara SB, Khoury RH, Suss MM, Lecain AE, Crowley Jr WF 1996 Inhibin B secretion in males with...