Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Determinants of Growth Hormone Stimulation Test Accuracy in a Large Pediatric Cohort

Bashar Karmi
Medicare Labs
Maryam Karmi
Medicare Labs
Ibrahim Salhi
gmd@medicare.ps

Medicare Labs

Research Article

Keywords: growth hormone deficiency, GH stimulation testing, clonidine, glucagon, arginine, BMI, pediatric endocrinology, diagnostic accuracy, sampling duration

Posted Date: June 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9701919/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background: GH stimulation tests are widely used to diagnose GH deficiency (GHD), but test accuracy is substantially influenced by agent pharmacodynamics, sampling duration, and patient physiology. No large single-center study from the Eastern Mediterranean has evaluated these interactions systematically.

Objective: To evaluate how stimulation agent, sampling duration, sex, age, and BMI influence GH stimulation test results and GHD diagnosis in a large pediatric cohort over a twenty-five-year period.

Methods: This retrospective audit included 1,487 patients, 795 males and 692 females, who underwent 1,542 pharmacological GH stimulation procedures at Medicare Laboratories, Palestine, over a twenty-five-year period. A total of 9,755 GH measurements were analyzed. The stimulation agents used were clonidine (n=757), glucagon (n=566), arginine (n=187), and levodopa (n=32). The primary outcome was the maximal stimulated peak GH concentration. Associations between GH response and anthropometric variables, including age, sex, and BMI, were assessed. Nonparametric statistical tests were used, and statistical significance was set at $p<0.05$.

Results: Glucagon and clonidine achieved normal GH responses (≥10 µg/L) in 50.2% and 46.1% of tests, vs. 18.7% for arginine and 9.4% for levodopa. Glucagon peak occurred at a mean of 123±35 min (median 120 min), that is beyond the conventional 120-minute endpoint. BMI showed a strong inverse correlation with peak GH (Spearman $\rho=-0.321$, $p<0.001$); normal response rates fell from 52.9% in underweight to 26.9–28.6% in overweight/obese patients. Females achieved modestly higher peak GH than males (10.79 vs. 10.02 µg/L, $p=0.007$), though peak timing did not differ by sex ($p=0.293$). Age group differences were significant (Kruskal–Wallis $p=0.009$), driven mainly by markedly lower responses in young adults over 18 years.

Conclusion: Glucagon protocols require extended sampling to 150–180 min. Uniform GH cutoffs without BMI adjustment risk misclassifying physiological obesity-related GH suppression as pathological deficiency. Agent-specific protocols and BMI-aware interpretation frameworks are essential for accurate GHD diagnosis.

1. Introduction

The diagnosis of growth hormone deficiency (GHD) in children remains one of the most clinically challenging areas in pediatric endocrinology [1]. Growth hormone is secreted in discrete, irregular pulses driven by the interplay of GH-releasing hormone (GHRH) and somatostatin, making random single measurements diagnostically uninformative [2]. Clinicians therefore rely on pharmacological stimulation tests, which use secretagogue agents to provoke a GH surge and assess pituitary reserve. Despite decades of widespread use, significant heterogeneity persists in the choice of stimulants, sampling duration, biochemical cutoff thresholds, and the handling of physiological confounders [3].

Internationally, GHD prevalence in children is estimated to be between 1 in 4,000 and 1 in 10,000, though reported rates vary substantially with the diagnostic criteria applied [4]. A core controversy surrounds the use of fixed biochemical cutoffs—most commonly 10 μg/L, and increasingly 7 μg/L in North American guidelines—applied uniformly without adjustment for patient characteristics [5]. Growing evidence indicates that GH responses exist along a continuous biological distribution rather than a binary pass–fail boundary, and that small changes in selected cutoffs can dramatically alter the proportion of children classified as GH-deficient [6].

A first major source of diagnostic uncertainty is the duration of blood sampling. For operational efficiency, many centers limit testing to 120 minutes. While this is adequate for fast-acting agents such as clonidine (peak typically at 60 min) and arginine (peak at 30–60 min), it may fail to capture delayed peak responses characteristic of slower-acting secretagogues. Glucagon in particular is known to act via indirect hypothalamic mechanisms with a prolonged time to peak, yet many protocols continue to truncate sampling at 120 minutes, risking systematic underestimation of GH peak amplitude [7].

A second major confounder is body mass index (BMI). Obesity is a well-established suppressor of both spontaneous and pharmacologically stimulated GH secretion, mediated through multiple mechanisms including hyperinsulinemia, elevated free fatty acids, enhanced somatostatinergic tone, and impaired GHRH responsiveness [8]. Meta-analyses have consistently demonstrated a significant inverse relationship between BMI and stimulated GH peaks across stimulation agents, age groups, and pubertal stages [9]. Despite this evidence, current international guidelines provide limited practical guidance on interpreting GH stimulation results in overweight and obese children [9].

These concerns are especially pertinent in the Eastern Mediterranean region, where childhood overweight and obesity rates have risen substantially over recent decades [10]. In this context, the risk of misclassifying obesity-related physiological GH suppression as permanent pituitary disease—with its attendant burden of long-term treatment—is substantial. The present retrospective audit was therefore designed to systematically characterize GH stimulation test performance across four pharmacological agents in a large, real-world pediatric cohort, with specific attention to the modifying effects of sampling duration, BMI, gender, and age on peak GH responses and diagnostic classification.

1.1 Study Rationale

This study addresses two complementary questions that remain incompletely answered in the literature: (1) How do the four most commonly used stimulation agents differ in their efficacy, temporal kinetics, and susceptibility to physiological modifiers? (2) To what extent do patient characteristics—particularly BMI, sex, and age—independently influence diagnostic classification? [9] Answering these questions in a large, consecutive cohort provides the empirical foundation for evidence-based protocol reform.

2. Methods

2.1 Study Design and Population

This study was a retrospective clinical audit of consecutive pediatric patients who underwent pharmacological GH stimulation testing between April 1, 2000, and March 31, 2026, at Medicare Laboratories, Palestine. The audit captured 1,487 unique patients contributing 1,542 stimulation procedures, generating a total of 9,755 individual GH measurements. Patients were included if they had been referred for evaluation of short stature or impaired growth velocity [11], had undergone at least one complete GH stimulation procedure, and had a complete test record. Biologically implausible values (e.g., GH > 200 μg/L) and incomplete records were excluded prior to analysis [12].

The cohort comprised 795 males (53.5%) and 692 females (46.5%), with a mean age of 10.20 ± 4.07 years (median 11.0 years, range 1–47 years). The age distribution was right-skewed, with the majority of patients in the 10–14 year (44.0%) and 5–10 year (34.0%) age groups. The relatively wide age range reflects the inclusion of a small number of young adults referred for transition-period evaluation. BMI data were available for 681 patients (45.8%); among these, the mean BMI was 17.23 ± 4.31 kg/m² (median 16.3 kg/m², range 11.7–38.4 kg/m²). The predominance of underweight patients (73.4% by standard BMI classification) reflects the growth-failure referral phenotype of the study population.

2.2 GH Stimulation Protocols

Four pharmacological stimulation agents were used, selected by the treating endocrinologist based on clinical indication, patient age, and institutional availability at the time of testing: clonidine (n = 757 procedures), glucagon (n = 566), arginine (n = 187), and levodopa (n = 32). Clonidine (0.15 mg/m² orally) and arginine (0.5 g/kg IV over 30 minutes) protocols used blood sampling at 0, 30, 60, 90, and 120 minutes. Glucagon protocols (0.03 mg/kg IM, maximum 1 mg) used blood sampling at 0, 30, 60, 90, 120, 150, and 180 minutes to allow capture of the delayed GH response [7]. Levodopa (100–500 mg orally based on weight) was sampled at 0, 30, 60, 90, and 120 minutes. GH concentrations were measured using locally validated immunoassay platforms specifically utilizing Tosoh, Mindray, and Cobas analyzer systems [12].

2.3 Outcome Measures

The primary outcome was the maximum stimulated GH peak concentration (μg/L) achieved at any sampling time point during each procedure. Secondary outcomes included: time to peak GH (minutes); time to reach GH ≥ 5 μg/L and ≥ 10 μg/L thresholds; response category classification: Insufficient (< 5 μg/L), Partial (5–10 μg/L), or Normal (≥ 10 μg/L) [5]; and the proportion of patients achieving a normal response per drug, per BMI category, per age group, and by sex.

2.4 Statistical Analysis

Non-normality of GH distributions was confirmed by the D'Agostino K^2 test (all distributions $p < 0.001$). Accordingly, all comparisons used nonparametric methods. Two-group comparisons (e.g., males vs. females) used the Mann–Whitney U test. Multi-group comparisons (stimulant agents, age groups, BMI

categories) used the Kruskal–Wallis H test followed by post-hoc pairwise Mann–Whitney U tests. Associations between continuous variables (BMI vs. peak GH) were quantified by Spearman's rank correlation coefficient (ρ) and Pearson's r for reference. Categorical outcome comparisons (response category distributions) used chi-square tests. Effect sizes for pairwise comparisons were expressed as rank-biserial correlation r. All tests were two-sided; α = 0.05.

2.5 Ethical Considerations

This study was conducted as a retrospective audit of anonymized clinical data collected during routine care. No direct patient contact or additional investigations were performed. All data were handled in accordance with institutional data governance and confidentiality standards. Formal ethics committee review was not required per institutional policy for de-identified retrospective audit studies.

3. Results

3.1 Study Population Overview

The complete cohort of 1,487 patients contributed 1,542 GH stimulation procedures. The overall median peak GH response across all procedures was 9.04 µg/L [IQR 5.20–13.99]. In terms of response classification: 671 procedures (43.5%) yielded a normal response (≥10 µg/L), 507 (32.9%) a partial response (5–10 µg/L), and 364 (23.6%) an insufficient response (<5 µg/L). The mean ± SD peak GH overall was 10.38 ± 7.87 µg/L (95% CI 9.99–10.77). Among the 681 patients with available BMI data, 73.4% were underweight by standard classification, 22.3% were in the normal range, 3.2% were overweight, and 1.0% were obese, reflecting the growth-failure referral phenotype.

3.2 Stimulant-Specific GH Response and Kinetics

Stimulant agent had a profound and highly significant effect on both peak GH magnitude (Kruskal–Wallis H=128.02, p<0.001) and temporal kinetics (H=842.72, p<0.001). Clonidine (n=757) and glucagon (n=566) were the most effective agents, achieving normal GH responses (≥10 µg/L) in 46.1% and 50.2% of tests, respectively. Arginine (n=187) yielded a normal response in only 18.7%, and levodopa (n=32) in 9.4%, the lowest of all four agents. Median peak GH values were 9.49 µg/L [5.71–14.29] for clonidine and 10.11 µg/L [6.27–14.90] for glucagon, versus 5.10 µg/L [2.53–8.30] for arginine and 4.77 µg/L [0.84–7.21] for levodopa. Summary statistics are presented in Table 1.

The temporal profiles differed dramatically between agents. Arginine produced the earliest peak (mean 44±32 min, median 30 min), levodopa and clonidine showed intermediate kinetics (means 50 and 66 min respectively, median 60 min for both), while glucagon was markedly delayed with a mean peak time of 123±35 min and median 120 min. Critically, for glucagon: mean time to reach GH ≥5 µg/L was 102±44 min, and mean time to GH ≥10 µg/L was 117±35 min—confirming that the majority of valid glucagon peaks occur at or beyond 120 minutes. Pairwise comparisons confirmed that clonidine and glucagon

significantly outperformed arginine and levodopa (all p<0.001), while the difference between clonidine and glucagon was modest (p=0.042). Arginine and levodopa did not differ significantly (p=0.148). Full pairwise statistics are presented in Table 5. Time-course GH response patterns by stimulant agent and gender are illustrated in Figure 1.

Table 1. Summary of GH Stimulation Test Results by Drug

Stimulant	N	M/F	Peak GH (μg/L) Median [IQR]	Normal ≥10 (%)	Time to GH≥5 (min, mean±SD)	Time to GH≥10 (min, mean±SD)	Peak Time (min, mean±SD)
Arginine	187	105/82	5.1 [2.5–8.3]	18.7%	38±29	36±30	44±32
Clonidine	757	406/351	9.5 [5.7–14.3]	46.1%	52±22	58±19	66±23
Glucagon	566	303/263	10.1 [6.3–14.9]	50.2%	102±44	117±35	123±35
Levodopa	32	18/14	4.8 [0.8–7.2]	9.4%	42±15	40±17	48±18

IQR = interquartile range; SD = standard deviation; M = male; F = female. Peak GH and timing values represent per-procedure data.

3.3 Peak GH Distribution and Response Category Analysis

Figure 2 presents the peak GH distribution and response category breakdown across all four stimulant agents. The violin plot (Fig 2, left panel) illustrates the markedly right-skewed, highly variable nature of GH responses, with clonidine showing a notably wide distribution including extreme outliers up to 126.6 μg/L. The stacked bar chart (Fig 2, center) visually confirms the marked superiority of clonidine and glucagon in achieving normal responses, with over 50% of glucagon tests and ~46% of clonidine tests exceeding the ≥10 μg/L threshold. Arginine and levodopa are dominated by insufficient partial responses. The boxplot stratified by sex (Fig 2, right panel) demonstrates that females tend toward slightly higher peak values across all agents, with overlapping distributions confirming that the sex effect, while statistically significant overall, has limited practical magnitude.

3.4 Gender Analysis

Overall, females achieved a modestly but significantly higher mean peak GH than males (10.79±7.42 vs. 10.02±8.23 μg/L; Mann–Whitney U=271,832, p=0.007). Response category distribution also differed by sex (χ^2=7.39, p=0.025), with females more likely to achieve a normal response (46.2% vs. 41.2%) and

less likely to have an insufficient response (20.6% vs. 26.2%). However, peak GH timing was identical between sexes (mean 85.6 vs. 82.9 min; p=0.293), as was time to GH ≥5 µg/L (70.9 vs. 70.5 min; p=0.977). BMI did not differ by sex (p=0.789). Drug-specific analysis showed that the female advantage was significant for arginine (7.14 vs. 5.48 µg/L, p=0.004) and glucagon (12.25 vs. 10.83 µg/L, p=0.026), but not for clonidine (p=0.763) or levodopa (p=0.196). Full gender comparison statistics are presented in Table 2.

Table 2. Gender Comparison of GH Stimulation Outcomes (Mann–Whitney U Test)

Variable	Male (N=832)	Female (N=710)	p-value	Test
Age (years, mean±SD)	10.69±4.08	9.65±3.94	0.0000***	MWU
Height (cm, mean±SD)	129.3±18.8	124.7±16.3	0.0000***	MWU
BMI (kg/m², mean±SD)	17.18±3.61	17.31±4.94	0.789 ns	MWU
Peak GH (µg/L, mean±SD)	10.02±8.23	10.79±7.42	0.007**	MWU
Peak time (min, mean±SD)	82.9±40.7	85.6±42.9	0.293 ns	MWU
Time to GH≥5 (min)	70.5±40.8	70.9±43.6	0.977 ns	MWU
Normal response ≥10 µg/L	343 (41.2%)	328 (46.2%)	0.025*	χ^2
Insufficient response <5 µg/L	218 (26.2%)	146 (20.6%)	0.025*	χ^2

MWU = Mann–Whitney U test; χ^2 = chi-square test; ns = not significant; * p<0.05; ** p<0.01; *** p<0.001.

3.5 Age Group Analysis

Age group had a statistically significant effect on peak GH (Kruskal–Wallis H=13.55, p=0.009). Within the pediatric age range, responses were broadly comparable: mean peak GH values of 10.98 µg/L (<5y), 10.63 µg/L (5–10y), 10.07 µg/L (10–14y), and 10.56 µg/L (14–18y) showed no clinically meaningful trend. The exception was the small adult group (>18y, n=10), with a mean peak GH of only 4.17±4.89 µg/L and an 80.0% insufficient response rate, consistent with the physiologically lower GH secretory capacity in early adulthood and the age-related increase in BMI (mean BMI 20.14 kg/m²). Normal response rates were highest in the youngest group (<5y: 47.6%) and the adolescent group (14–18y: 47.0%). BMI increased progressively across age groups (Kruskal–Wallis p<0.001), from 15.16±2.01 kg/m² in children <5y to 20.14±4.39 kg/m² in those >18y. Age-group-specific GH stimulation outcomes are summarized in Table 3.

Table 3. GH Stimulation Outcomes by Age Group

Age Group	N	Peak GH (µg/L, mean±SD)	Normal ≥10 (%)	Partial 5–10 (%)	Insuf. <5 (%)	Mean BMI (kg/m²)	Mean Peak Time (min)
<5 years	210	10.98±7.07	47.6%	33.8%	18.6%	15.16±2.01	79.9±40.4
5–10 years	530	10.63±9.19	45.3%	33.6%	21.1%	16.22±3.09	80.7±39.7
10–14 years	675	10.07±6.93	41.3%	34.2%	24.4%	18.12±4.97	87.6±43.1
14–18 years	115	10.56±7.93	47.0%	27.0%	26.1%	19.87±4.34	87.4±44.7
>18 years	10	4.17±4.89	20.0%	0.0%	80.0%	20.14±4.39	81.0±34.8

Kruskal–Wallis H=13.55, p=0.009 for peak GH; H=106.41, p<0.001 for BMI across age groups.

3.6 BMI Analysis and GH Suppression

Among 681 patients with available BMI data (mean 17.23±4.31 kg/m²), Spearman correlation analysis revealed a strong and highly significant inverse association between BMI and peak GH (ρ=−0.321, p<0.001; Pearson r=−0.231, p<0.001). As the BMI category increased, mean peak GH concentration and normal response rates fell substantially and monotonically. Underweight patients (n=537, 78.9% of those with BMI data) achieved a mean peak GH of 11.96±9.14 µg/L with a 52.9% normal response rate. In stark contrast, patients with normal BMI (n=111) had a mean peak GH of 7.64±5.07 µg/L (normal response rate 26.9%), and overweight/obese patients (n=33 combined) showed mean values of 5.96–6.04 µg/L with normal response rates of 27–29%. These data demonstrate that moving from underweight to any higher BMI category is associated with a halved normal response rate.

Figure 3 illustrates the BMI–GH relationship. The scatter plot confirms the inverse linear trend (r=−0.231, p<0.001), with the relationship clearly visible despite marked individual variability at lower BMI values. The BMI distribution in this cohort (left panel, N=681, mean 17.2 kg/m²) and the peak GH by BMI category boxplot (right panel) consistently reinforce the pattern of GH suppression with increasing adiposity. Complete BMI category statistics are presented in Table 4.

Table 4. GH Stimulation Outcomes by BMI Category

BMI Category	N	% of total	Mean Peak GH (µg/L, mean±SD)	95% CI	Normal Response %	p vs UW
Underweight (<18.5)	537	78.9%	11.96±9.14	11.19–12.73	**52.9%**	ref
Normal (18.5–25)	111	16.3%	7.64±5.07	6.85–8.43	26.9%	<0.001
Overweight (25–30)	26	3.8%	5.96±4.61	4.10–7.83	26.9%	<0.001
Obese (>30)	7	1.0%	6.04±6.32	0.19–11.88	28.6%	<0.001

Spearman ρ=−0.321, p<0.001; Pearson r=−0.231, p<0.001 for BMI vs. peak GH. UW = underweight reference group. N=681 with BMI data (45.8% of full cohort); underweight n=537, normal n=111, overweight n=26, obese n=7.

Table 5. Pairwise Drug Comparisons: Mann–Whitney U Tests for Peak GH

Comparison	U Statistic	p-value	Significance	Effect size r
Arginine vs. Clonidine	40,584	<0.001	***	large
Arginine vs. Glucagon	27,328	<0.001	***	large
Arginine vs. Levodopa	3,472	0.148 ns	ns	0.098 (small)
Clonidine vs. Glucagon	200,276	0.042	*	0.056 (small)
Clonidine vs. Levodopa	18,664	<0.001	***	0.185 (small-mod)
Glucagon vs. Levodopa	14,390	<0.001	***	0.229 (moderate)

ns = not significant; * p<0.05; *** p<0.001. Effect size r = rank-biserial correlation. 'large' indicates effect size could not be calculated due to complete or near-complete separation.

4. Discussion

This large, single-center retrospective audit of 1,487 pediatric patients undergoing 1,542 GH stimulation procedures provides a comprehensive characterization of real-world GH stimulation testing performance. The data demonstrate that test outcomes are shaped not only by pituitary reserve, but substantially by three modifiable or clinically important factors: the choice of pharmacological agent, the duration of blood sampling, and patient BMI. These findings have direct implications for how GH stimulation protocols are designed and how results are interpreted.

4.1 Stimulant Efficacy and the Case for Agent-Specific Protocols

The most prominent finding is the marked hierarchy in stimulant efficacy [13]. Glucagon and clonidine achieved normal GH responses in approximately half of all procedures, while arginine (18.7%) and levodopa (9.4%) underperformed substantially. The statistically significant but clinically modest difference between clonidine and glucagon ($p=0.042$, effect size $r=0.056$) suggests these two agents are broadly comparable in efficacy, though their temporal profiles differ fundamentally.

The delayed pharmacodynamics of glucagon deserve particular emphasis. With a mean peak time of 123±35 minutes and a median peak at 120 minutes—and with the mean time to reach the ≥10 µg/L threshold at 117±35 minutes—a conventional 120-minute protocol will systematically truncate the rising phase of the glucagon response for a substantial proportion of patients. Any 120-minute cut-off will therefore produce a false-negative result rate that is difficult to estimate precisely in retrospective data, but our temporal data strongly imply that many delayed-peak patients would be misclassified as GH-deficient if testing were stopped at 120 minutes. This is consistent with published reports from other centers [7] and supports the recommendation that glucagon-based GH stimulation testing should routinely extend sampling to at least 150 minutes, and preferably 180 minutes [5].

The inferior performance of arginine and levodopa in this cohort—with over 48% and 56% of tests, respectively, yielding insufficient responses—raises legitimate questions about their continued use as first-line diagnostic tools. The absence of a statistically significant difference between arginine and levodopa ($p=0.148$), combined with their similarly low normal response rates, suggests that both agents may be less suitable than clonidine or glucagon for primary diagnostic testing in this population. Whether these findings reflect differences in patient characteristics [12], or true pharmacological inferiority requires prospective evaluation [14].

4.2 BMI as a Major Physiological Confounder

The strong, dose-dependent inverse relationship between BMI and peak GH responses is one of the most clinically actionable findings of this study. The fall in mean peak GH from 11.96 µg/L in underweight patients to 5.96–6.04 µg/L in overweight and obese patients, respectively—a greater than twofold reduction—and the fall in normal response rates from 52.9% to approximately 27%, illustrate the scale of BMI-mediated GH suppression in clinical practice [9].

This suppression reflects physiological modulation of the somatotropic axis rather than structural pituitary disease. Hyperinsulinemia in overweight individuals reduces GHRH responsiveness and enhances somatostatin tone [8]; elevated free fatty acids further blunt GH secretion at the pituitary level [15]. Critically, this suppression is reversible with weight normalization [9], meaning that it represents a potentially transient, non-pathological state. Classifying overweight or obese children as GH-deficient on

the basis of a suppressed stimulation test—without accounting for their elevated BMI—therefore risks initiating years of daily recombinant GH injections for a condition that may not represent true pituitary pathology.

The practical implication is clear: diagnostic frameworks must incorporate BMI adjustment. Options include BMI-stratified normative GH response ranges [9], a requirement for weight normalization (or a minimum threshold of weight loss) before testing in overweight and obese children, or the use of GH-stimulating protocols combined with BMI-dependent correction factors. Until such frameworks are prospectively validated, the clinical minimum is to explicitly document BMI at the time of testing and to interpret GH results in that physiological context.

4.3 Gender Effects: Modest Magnitude, Limited Clinical Impact

While females achieved statistically higher peak GH than males in this cohort (p=0.007), the magnitude of this difference is modest (mean difference approximately 0.77 µg/L) and does not translate into a meaningful difference in peak timing (p=0.293) or time-to-threshold metrics (p=0.977). The difference in response category distribution (normal response: 46.2% females vs. 41.2% males, p=0.025) is statistically significant but represents a small absolute difference that is unlikely to justify sex-specific diagnostic cutoffs in the absence of BMI and pubertal stage control [1]. The drug-specific sex differences observed for arginine and glucagon but not clonidine or levodopa suggest that sex may interact with specific mechanisms of GH release in a pharmacologically selective manner, a hypothesis that merits a dedicated prospective study [13].

4.4 Age Effects and Transition-Age Patients

The broadly comparable GH responses across pediatric age groups (<5y through 14–18y) support the use of consistent diagnostic cutoffs within the pediatric range [11], at least in the absence of additional pubertal priming. The markedly lower responses in patients over 18 years (mean peak GH 4.17 µg/L, 80% insufficient responses) are consistent with the biology of declining GH secretion in early adulthood and increasing BMI [5], and underscore the importance of different diagnostic approaches for transition-age patients [11].

4.5 Toward Evidence-Based Protocol Reform

Collectively, these data provide a robust empirical foundation for two targeted and immediately implementable protocol reforms. First, sampling duration must be agent-specific: glucagon protocols should extend to 180 minutes as standard, with 150 minutes as an absolute minimum [7]. Second, GH stimulation results must be interpreted in the context of patient BMI, with either explicit documentation

and qualitative adjustment, or formally validated BMI-adjusted normative ranges [9]. Implementing these changes in high-prevalence obesity settings [10] is likely to meaningfully reduce the proportion of false-positive GHD diagnoses and the unnecessary initiation of recombinant GH therapy.

Limitations of this study include its retrospective, single-center design; the incompleteness of BMI data for a substantial proportion of patients; the lack of pubertal staging data, which is an important independent modifier of GH responses; and the absence of confirmatory second-stimulation test data, which is required by many guidelines before initiating treatment [5]. The levodopa group was substantially smaller than the other three, limiting the precision of comparative estimates. Prospective, multi-center studies with standardized protocols, complete anthropometric and pubertal staging data, and pre-specified BMI-adjusted cutoffs are needed to validate and extend these findings.

5. Conclusion

In this large retrospective cohort of 1,487 patients undergoing 1,542 GH stimulation procedures, glucagon and clonidine were the most effective pharmacological agents, each achieving normal GH responses in approximately half of all tests. Glucagon exhibited markedly delayed peak kinetics (mean peak time 123±35 min, time to GH ≥10 μg/L 117±35 min), confirming that conventional 120-minute sampling protocols are insufficient for this agent [7] and will generate false-negative GH peaks in a clinically significant proportion of patients. BMI exerted a strong, independent, and dose-dependent inverse effect on stimulated GH secretion (Spearman $\rho=-0.321$, $p<0.001$), with normal response rates approximately halved across the transition from underweight to overweight/obese status. Females achieved modestly higher peak GH than males, though peak timing was similar, and age group differences were modest within the pediatric range.

These findings support two evidence-based changes to current clinical practice: (1) adoption of agent-specific sampling durations, with glucagon protocols extended to a minimum of 150–180 minutes [5]; and (2) integration of BMI into GH stimulation test interpretation [9], through formally validated BMI-adjusted reference ranges or at a minimum through explicit clinical acknowledgment of BMI-related suppression before initiating GHD treatment. These reforms are particularly important in clinical settings with high childhood obesity prevalence [10], where uniform cutoff application without BMI context risks systematic over-diagnosis of GH deficiency and unnecessary long-term treatment [9].

Declarations

Ethics Approval and Consent to Participate:

This study was approved by the Institutional Review Board (IRB) of Medicare Laboratories, Palestine (approval waiver granted under retrospective de-identified data exemption). The laboratory tests included in this study were performed as part of routine physician-requested clinical investigations. Data used in this retrospective cohort analysis were extracted from the laboratory information system (LIS) in

fully anonymized form, and no identifiable patient information was accessed or released. Due to the retrospective nature of the study and the use of fully de-identified data, the requirement for individual informed consent was waived by the IRB of Medicare Laboratories in accordance with institutional policy and applicable ethical standards for retrospective audit studies.

Clinical Trial Number:

not applicable.

Consent for Publication:

Not applicable.

Conflict of Interest:

The authors declare no conflicts of interest relevant to this work.

Funding:

This research received no external funding.

Data Availability:

De-identified data are available from the corresponding author upon reasonable request, subject to institutional data governance approval.

Acknowledgements:

The authors thank the clinical and laboratory staff of Medicare Laboratories for their contributions to patient care and data collection over the audit period.

Author Contributions:

B.K. contributed to clinical supervision of patients, oversight of all stimulation testing procedures, and review of laboratory results. M.K. contributed to clinical laboratory work, including specimen processing and GH immunoassay measurements. I.S. contributed to manuscript preparation, data analysis, and statistical interpretation. All authors reviewed and approved the final manuscript.

References

1. Murray PG, Dattani MT, Clayton PE. Controversies in the diagnosis and management of growth hormone deficiency in childhood and adolescence. Arch Dis Child. 2016;101(1):96–100. doi:10.1136/archdischild-2014-307228
2. Giustina A, Veldhuis JD. Pathophysiology of the neuroregulation of growth hormone secretion in experimental animals and the human. Endocr Rev. 1998;19(6):717–797. doi:10.1210/edrv.19.6.0353
3. Yuen KCJ, Johannsson G, Ho KKY, Miller BS, Bergada I, Rogol AD. Diagnosis and testing for growth hormone deficiency across the ages: a global view of the accuracy, caveats, and cut-offs for diagnosis. Endocr Connect. 2023;12(7):e220504. doi:10.1530/EC-22-0504
4. Ibba A, Loche S. Diagnosis of GH deficiency without GH stimulation tests. Front Endocrinol (Lausanne). 2022;13:853290. doi:10.3389/fendo.2022.853290
5. Yuen KCJ, Biller BMK, Radovick S, et al. American Association of Clinical Endocrinologists and American College of Endocrinology guidelines for management of growth hormone deficiency in adults and patients transitioning from pediatric to adult care. Endocr Pract. 2019;25(11):1191–1232. doi:10.4158/GL-2019-0405
6. Binder G, Reinehr T, Ibáñez L, et al. GHD diagnostics in Europe and the US: an audit of national guidelines and practice. Horm Res Paediatr. 2020;93(7–8):404–415. doi:10.1159/000509432
7. Yuen KCJ, Tritos NA, Engelman EG, et al. Glucagon stimulation testing in assessing for adult growth hormone deficiency: current status and future perspectives. Pituitary. 2012;15(1):105–117. doi:10.1007/s11102-011-0348-8
8. Stanley TL, Grinspoon SK. Effects of growth hormone-releasing hormone on visceral fat, metabolic, and cardiovascular indices in human studies. Growth Horm IGF Res. 2015;25(2):59–65. doi:10.1016/j.ghir.2014.12.005
9. van den Akker ELT, Hokken-Koelega ACS, et al. Impact of body mass index on growth hormone stimulation tests in children and adolescents: a systematic review and meta-analysis. Crit Rev Clin Lab Sci. 2022;59(8):576–595. doi:10.1080/10408363.2021.1956423
10. Farrag NS, Cheskin LJ, Farag MK. A systematic review of childhood obesity in the Middle East and North Africa (MENA) region: prevalence and risk factors meta-analysis. Adv Pediatr Res. 2017;4:8. doi:10.12715/apr.2017.4.8
11. Collett-Solberg PF, Ambler G, Backeljauw PF, et al. Diagnosis, genetics, and therapy of short stature in children: a Growth Hormone Research Society international perspective. Horm Res Paediatr. 2019;92(1):1–14. doi:10.1159/000502231
12. Bidlingmaier M, Freda PU. Measurement of human growth hormone by immunoassays: current status, unsolved problems and clinical consequences. Growth Horm IGF Res. 2010;20(1):19–25. doi:10.1016/j.ghir.2009.09.005
13. Ghigo E, Bellone J, Aimaretti G, et al. Reliability of provocative tests to assess growth hormone secretory status. Study in 472 normally growing children. J Clin Endocrinol Metab.

1996;81(9):3323–3327. doi:10.1210/jcem.81.9.8784091

14. Collu R, Leboeuf G, Letarte J, Ducharme JR. Stimulation of growth hormone secretion by levodopa-propranolol in children and adolescents. Pediatrics. 1975;56(2):262–266.
15. Misra M, Bredella MA, Tsai P, Mendes N, Miller KK, Klibanski A. Lower growth hormone and higher cortisol are associated with greater visceral adiposity, intramyocellular lipids, and insulin resistance in overweight girls. Am J Physiol Endocrinol Metab. 2008;295(2):E385–E392. doi:10.1152/ajpendo.00052.2008

Figures

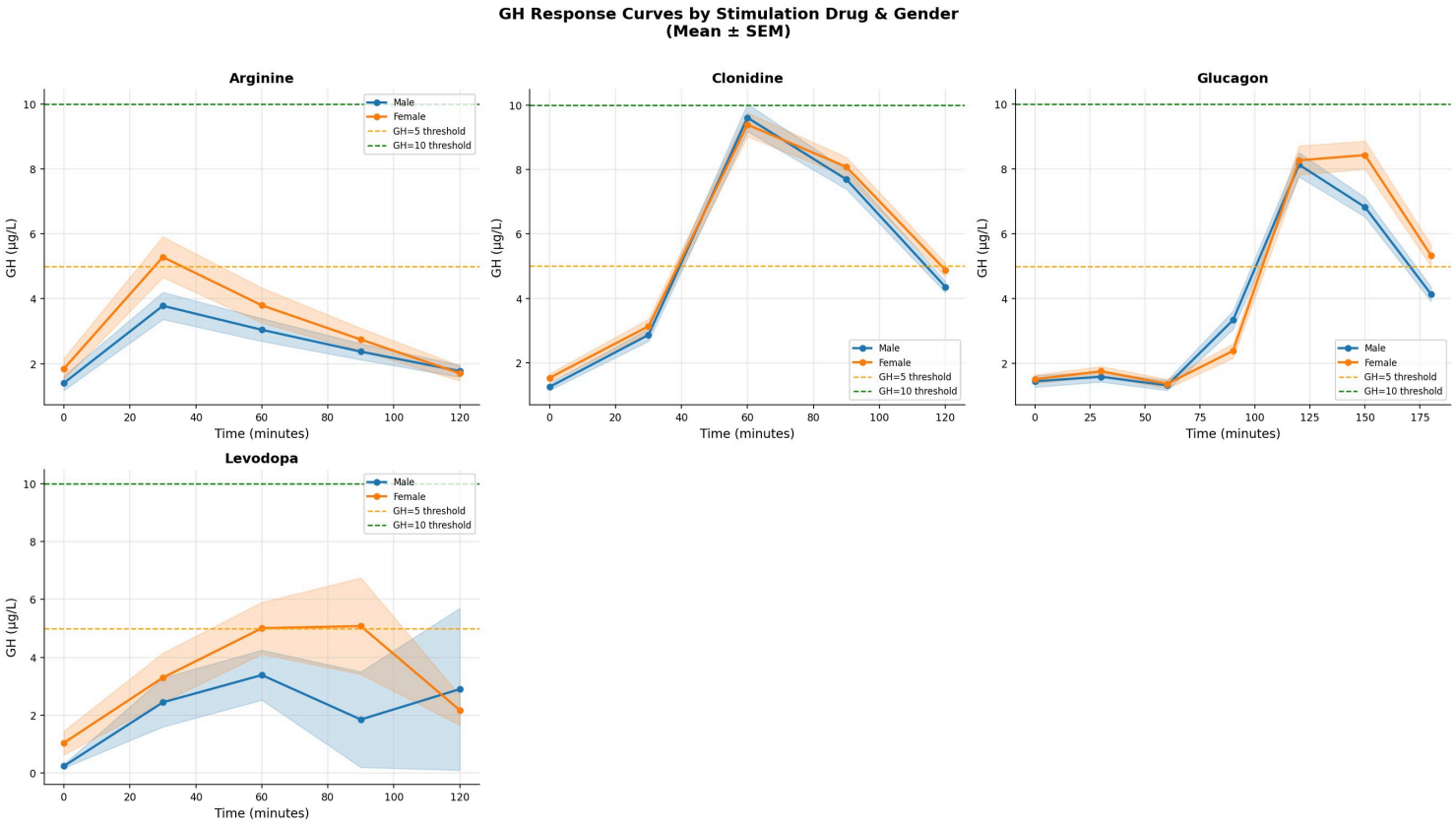


Figure 1

GH Response Curves by Stimulation Drug and Gender (Mean ± SEM)

Figure 1. Time-course GH response curves (mean ± SEM) for each stimulant agent, stratified by sex. Dashed lines indicate the GH=5 µg/L (yellow) and GH=10 µg/L (green) diagnostic thresholds. Note the markedly extended peak plateau for glucagon (≥120 min) compared to the earlier peaks for clonidine (60 min) and arginine (30 min).

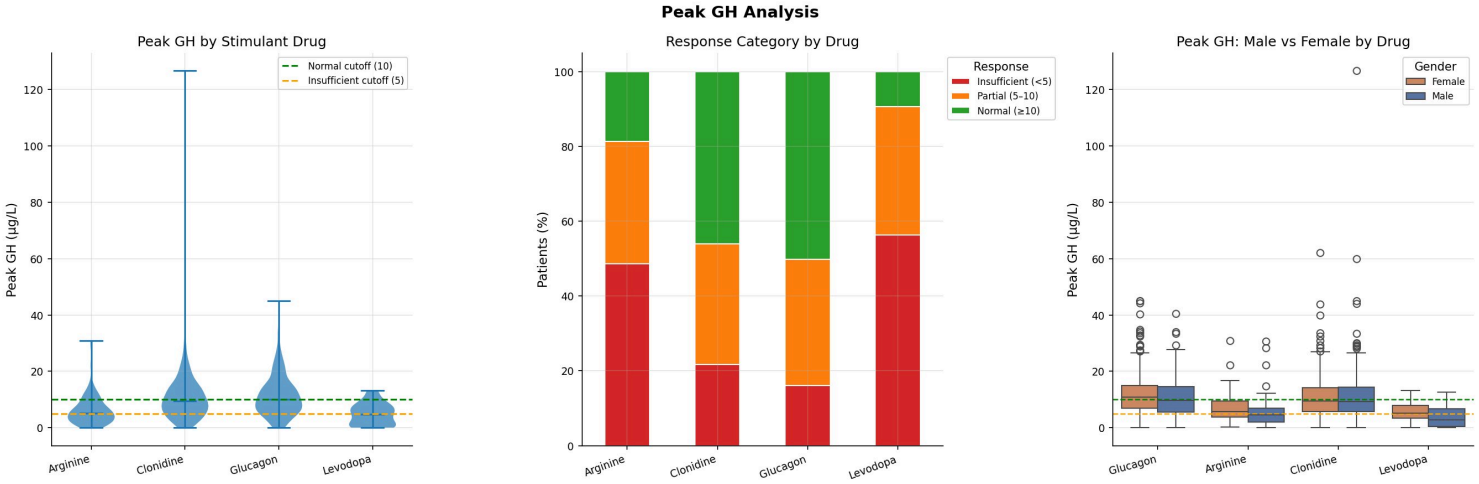


Figure 2

Peak GH Analysis: Distribution, Response Categories, and Gender Comparison

Figure 2. Left: Violin plots of peak GH distribution by stimulant; yellow/green dashed lines indicate 5 and 10 µg/L thresholds. Center: Stacked bar charts showing proportions of insufficient (<5), partial (5–10), and normal (≥10 µg/L) responses per agent. Right: Box plots of peak GH by sex for each agent.

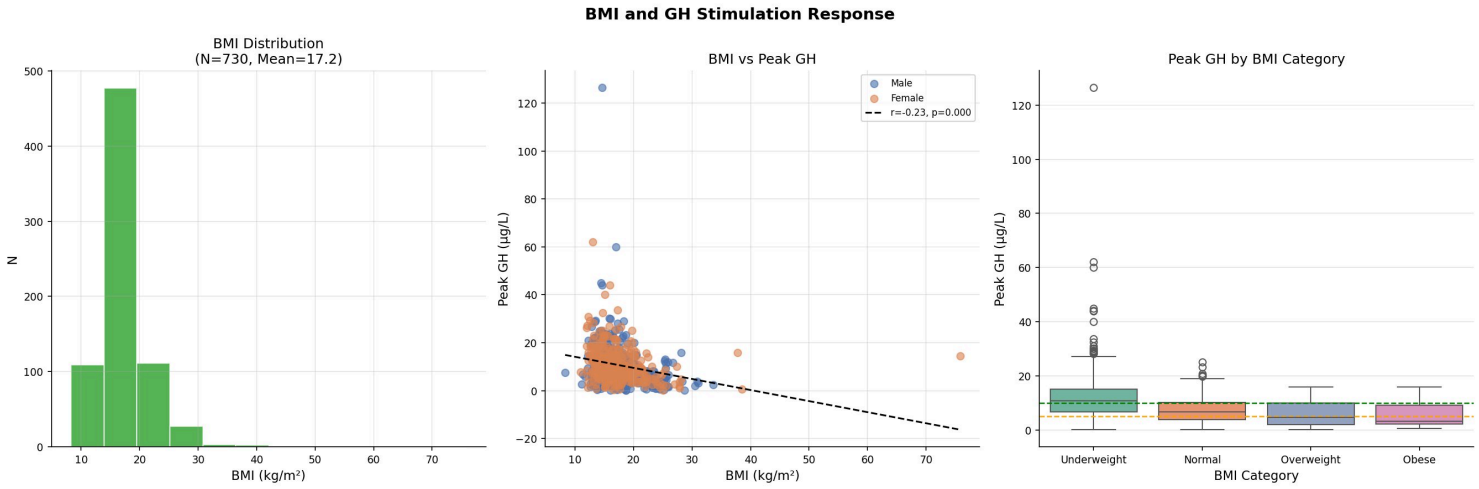


Figure 3

BMI and GH Stimulation Response

Figure 3. Left: BMI distribution of the cohort (N=681, mean 17.2 kg/m²). Center: Scatter plot of BMI vs. peak GH by sex; dashed line represents linear regression (r=−0.23, p<0.001). Right: Box plots of peak GH by BMI category. Yellow/green dashed lines indicate 5 and 10 µg/L thresholds.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableA.docx