Published in final edited form as:

*Horm Res Paediatr*. 2021 ; 94(9-10): 353–363. doi:10.1159/000520250.

# Growth hormone stimulation testing patterns contribute to sex differences in pediatric growth hormone treatment

**Camilia Kamoun**[a], **Colin Patrick Hawkes**[a,b,c], **Hareesh Gunturi**[d], **Andrew Dauber**[e,f], **Joel N Hirschhorn**[g,h], **Adda Grimberg**[a,b,i]

[a]Division of Endocrinology and Diabetes, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[b]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[c]Department of Paediatrics and Child Health, University College Cork, Cork, Ireland

[d]Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[e]Division of Endocrinology, Children's National Hospital, Washington, District of Columbia, USA

[f]Department of Pediatrics George Washington University School of Medicine and Health Sciences, Washington, District of Columbia, USA

[g]Division of Endocrinology, Boston Children's Hospital, Boston, MA, USA

[h]Departments of Pediatrics and Genetics, Harvard Medical School, Boston, MA, USA

[i]Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, USA

## Abstract

**Introduction:** Males are twice as likely as females to receive pediatric growth hormone (GH) treatment in the United States, despite similar distributions of height z-scores in both sexes. Male predominance in evaluation and subspecialty referral for short stature contributes to this observation. This study investigates whether sex differences in GH stimulation testing and subsequent GH prescription further contribute to male predominance in GH treatment.

**Methods:** Retrospective chart review was conducted of all individuals, age 2–16 years, evaluated for short stature or poor growth at a single large tertiary referral center between 2012–2019. Multiple logistic regression models were constructed to analyze sex differences.

**Results:** Of 10,125 children referred for evaluation, a smaller proportion were female (35%). More males (13.1%) than females (10.6%) underwent GH stimulation testing ($p<0.001$) and did so at heights closer to average (median height z-score −2.2 [interquartile range (IQR) −2.6, −1.8] vs. −2.5 [IQR −3.0, −2.0], respectively; $p<0.001$). The proportion of GH prescriptions by sex

Corresponding Author: Camilia Kamoun, MD, Division of Endocrinology and Diabetes, Children's Hospital of Philadelphia, 3500 Civic Center Blvd, Philadelphia, PA, 19104, Tel: 01-215-590-3174, kamounc@chop.edu.

Author Contributions

A.G., C.P.H., A.D., and J.H. conceived the database used for the study. C.K., A.G., and C.P.H. designed the study. H.G. extracted data from the EHR. C.K. performed data analysis and interpretation, which A.G. and C.P.H. reviewed. C.K. drafted the manuscript. All authors reviewed and edited the manuscript.

was similar by stimulated peak GH level. Predictor variables in regression modeling differed by sex: commercial insurance predicted GH stimulation testing and GH prescription for males only, whereas lower height z-score predicted GH prescription for females only.

**Conclusions:** Sex differences in rates of GH stimulation testing, but not subsequent GH prescription based on response to GH stimulation testing, seem to contribute to male predominance in pediatric GH treatment. That height z-score predicted GH prescription in females but not males raises questions about the extent to which sex bias—from children, parents and/or physicians—, as opposed to objective growth data, influence medical decision-making in the evaluation and treatment of short stature.

### Keywords

Growth hormone stimulation testing; sex; sex difference; growth hormone treatment

## Introduction

Short stature is associated with heightened concern when present in boys compared to girls in primary care settings [1, 2]. Yet, in a population of children referred to endocrinologists for evaluation of short stature, 41% of girls, compared to only 15% of boys, were found to have organic disease ($p < 0.0001$) [1]. Further, despite absence of sex differences in distribution of height z-scores in a large pediatric primary care population in the United States of America (U.S.) [3], boys outnumbered girls 2:1 for all growth hormone (GH) indications and 3:1 for idiopathic short stature (ISS) in four U.S. pediatric GH registries [4, 3]. The burden of GH treatment includes expense—$35,000-$52,000 per inch of height gained [5, 6] with a mean of $1099 per year paid by the commercially insured patient in 2016 [7]—, potential side effects [8], and daily injections [9].

Various steps in the process from initial concern about growth to GH treatment have been associated with a sex disparity. Males are more likely to be screened by primary care clinicians for GH deficiency (GHD) [2] and to be referred to a pediatric endocrinologist for evaluation of short stature [1]. Pediatric endocrinologists are also more likely to prescribe GH treatment to boys than girls, when this was assessed in theoretical cases of short non-GHD children [10]. In a survey of parents of pediatric primary care patients, the acceptable height cut-off regarding short stature was higher for male versus female heights [11].

As a key step in the decision to treat with GH, GH stimulation testing also may contribute to sex disparities in GH treatment. Racial and ethnic differences in the likelihood of undergoing GH stimulation testing recently have been demonstrated [12]. The significant imprecision of GH stimulation testing leaves room for potential sex bias in ordering and interpreting the test as well. GH stimulation testing has poor sensitivity and specificity [13]. Additionally, the normal threshold for peak stimulated GH concentration is not well defined, because so many test and patient factors contribute to peak stimulated GH concentrations [13]. While the most recent international consensus paper suggests using a peak GH level of < 7 ng/mL to define GHD, this cutoff used to be 10 ng/mL [14–16]. As such, a level in the range of 7–10 ng/mL has become a grey zone where some clinicians may diagnose partial GHD whereas others would not, leaving more room for influence by subjective factors. This study sought to

examine whether sex differences in ordering and interpreting GH stimulation tests contribute to disparities in GH treatment by testing the following hypotheses: 1) A greater proportion of males than females referred for evaluation of short stature or poor growth undergo GH stimulation testing and do so at lesser height deficits; and 2) Males with borderline peak GH levels (between 7–10 ng/mL) are more likely than females to be prescribed GH treatment. The study also sought to explore whether sex differences exist in clinical and demographic variables associated with GH stimulation testing and GH prescription.

## Materials and Methods

### Study design and setting

A retrospective cohort study was performed of children undergoing initial evaluation of short stature or poor growth between January 1, 2012 and December 31, 2019 at a single large tertiary pediatric endocrinology referral center and its satellite offices. This timeframe was chosen because the GH assay used for measuring GH as part of the stimulation testing remained consistent throughout this period. The Institutional Review Board of the Children's Hospital of Philadelphia reviewed the study and determined it met the exemption criteria per 45 CFR 46.104(d) 4(iii). A waiver of HIPAA authorization per 45 CFR 164.512(i)(2)(ii) was granted for accessing identifiable information from the electronic health records (EHR).

### Study population

Children aged 2 to 16 years were included if they were a new patient with a visit diagnosis of short stature or poor growth. Specifically, individuals with clinical diagnoses that included the words "growth," "short," "small," "stature," and "height" were extracted and those with diagnoses relevant to short stature or poor growth were included. Children previously diagnosed with GHD or who were prescribed GH without preceding stimulation testing were excluded.

### Growth hormone stimulation testing

As per clinical routine at the study center, GH stimulation testing included two GH stimulation tests performed sequentially, starting at 08:00 in the morning after a fast (from midnight for children ≥ 3 years old, and for 2 hours for those less than age 3). At the start of the test, an intravenous catheter was placed for serial blood sampling. Provocative testing with clonidine and arginine was commonly performed, while glucagon was rarely used in place of clonidine. Oral clonidine was dosed as follows: 0.15 mg for patient weight >50 kg, 0.1 mg for weight 15–50 kg, and 0.15 mg/m2 for weight <15 kg. Following clonidine administration, GH concentrations were measured at baseline, 30, 60, 90 and 120 minutes. After this was completed, 0.5 g/kg (max 30 g) of arginine was administered intravenously over 30 minutes, and GH concentrations were measured at baseline, 15, 30, 45, 60 and 90 minutes. Where glucagon was used as the GH stimulus, 1 mg was administered intravenously and GH concentrations were measured at baseline, 90, 120, 150 and 180 minutes. All GH concentrations were measured clinically using the Siemens Immulite assay (Siemens Medical Solutions USA, Inc, Malvern, PA), standardized to IS

98/574, as recommended by the 2011 Consensus Statement [17] and Pediatric Endocrine Society GH Guidelines [16].

## Data extraction

Demographic and clinical data from the first endocrinology clinic evaluation were extracted. As per the endocrinology clinic's routine, visit height measurements were the average of triplicate measurements made with a wall-mounted stadiometer. Height z-scores based on the Centers for Disease Control and Prevention (CDC) growth charts were derived to standardize height for age and sex [18]. BMI similarly was extracted as z-score based on CDC growth charts. If a height measurement was available from a non-endocrinology encounter that occurred prior to the initial endocrinology evaluation, this information was also extracted, along with the corresponding height z-score, and used to calculate annualized height velocity and annualized change in height z-score, respectively.

Pubertal status, defined by Tanner staging, was obtained from the clinical exam documented by the endocrinologist. Tanner staging for females was based on breast development, with assignment of half stages for those whose Tanner staging was recorded as between two stages. Testicular volume was used to define Tanner stage for males (1 = < 4 cc, 2 = 4–8 cc, 3 = 9–12 cc, 4 = 13–19 cc, 5 = 20–25 cc). Tanner 1 was classified as prepuberty, Tanner 1.5–3 as early puberty, Tanner 3.5–5 as late puberty. Parental heights were self-reported, with sex-adjusted mid-parental height (MPH) [19] transformed into z-scores relative to the CDC growth chart data [18]. Comparison to MPH was defined as height z-score minus MPH z-score. Race, ethnicity, and sex were determined by parental or child report. Insurance type as recorded in the EHR was categorized as commercial or government. Clinic location was categorized as urban or non-urban based on the U.S. Census Bureau's definition of population density for an urban location (1500 people per square kilometer) [20]. IGF-I measurements at or near the time of the visits (but before GH stimulation testing) were extracted as available.

For individuals who underwent GH stimulation testing, peak GH concentrations and sex hormone priming status were extracted, as well as whether pituitary MRI was obtained and whether pharmacological GH therapy was initiated. Provider preference determined sex hormone priming criteria and protocol. Peak GH values were categorized into three groups representing different cutoffs for GH treatment [14–16]: <7 mg/dL, 7–10 mg/dL, and >10 mg/dL.

## Statistical analysis

Descriptive statistics were used to present demographic and clinical data, with continuous variables reported as median (interquartile range [IQR]). Wilcoxon rank sum test was used to compare continuous variables. Categorical variables were compared using chi-squared test or, in the case of two categories within a categorical variable, a two-tailed two-sample test of proportions. Logistic regression analysis was conducted to determine the odds ratios (OR) of GH stimulation testing and GH prescription. Regression models were constructed to include variables that were predicted and found to be significant (p-value ≤0.05) on univariate analysis against the outcome of interest and to control for variables clinically

significant to the outcomes of interest. Multiple imputation was conducted to allow for inclusion in the models of subjects who had data missing for 1–2 variables. Only variables with less than 10% of observations missing were imputed and included in the full models. Separate regression models that included only individuals with IGF-I z-scores were also constructed to assess the effect of IGF-I z-scores. Logistic regression also was used to identify significant predictor variables by sex. Statistical analyses were conducted using Stata/MP© version 16.1 (StataCorp, College Station, Texas, USA).

## Results

### Endocrine clinic population

The study population included 10,125 children referred for evaluation of short stature or poor growth, of whom 3542 (35%) were female. Demographic and clinic characteristics by sex are shown in Table 1. In this population of patients receiving new growth evaluations, a greater proportion of females had government insurance and were in late puberty. The proportion of non-Hispanic Black (NHB) individuals was lower for females than males, but a greater proportion of females were Hispanic and Other (Table 1). Sex ratio differed by race-ethnicity grouping, as well (Table 1). Asian race, the largest group in the Other category, was greater for females (females 265 [7.5%], males 323 [4.9%], p-value <0.0001). Females were shorter on average by height z-score but not by comparison to MPH (Table 1). The calculated values of annualized height velocity and annualized change in height z-score prior to the clinic visit, available for only a limited number of participants, were statistically different between sexes but the difference was not clinically significant.

### GH stimulation test population

A total of 1,245 children underwent GH stimulation testing. Twenty-one children (females 2 [0.5%], males 19 [2.2%]) received sex hormone priming prior to testing. Females were younger (females 9.4 [7.4, 11.7] years, males 10.9 [8.0, 13.3] years, p-value <0.001). Females had lower median height Z-score compared to males (Fig. 1) and greater height deficit in comparison to MPH (females −1.9 [−2.5, −1.4] vs. males −1.8 [−2.3, −1.3], p-value 0.02). Females had higher IGF-I z-scores (females −1.4 [−2.1, −0.5] vs. males −1.7 [−2.2, −0.9], p-value 0.01), available for 68% of females and 70% of males (non-significant [N.S.]). There were no sex differences in pubertal status, annualized change in height z-score and growth velocity, insurance type, race-ethnicity grouping or clinic location.

A smaller proportion of referred females than males underwent GH stimulation testing (Fig. 1). Unadjusted odds ratio for undergoing stimulation testing for females compared to males was 0.78 (95% confidence interval [CI] 0.69–0.89; p-value <0.001). In univariate analysis, sex, height z-score, pubertal status, insurance type, clinic location, race-ethnicity grouping, comparison to MPH and IGF-I z-scores were significant predictors of GH stimulation testing (Table 3). Controlling for these variables in multivariate analysis, excluding IGF-I z-scores and comparison to MPH (due to >10% of values missing for these variables), females had a 0.67 (95% CI 0.59–0.77; p-value <0.001) odds of undergoing stimulation testing compared to males. Sub-analysis suggested different sex effects depending on race-ethnicity grouping, controlling for height z-score, pubertal status, insurance type, and clinic location, although

these differences did not reach statistical significance (Fig. 2). In multivariate analysis with IGF-I z-score and comparison to MPH (n=4757), females had a 0.79 (95% CI 0.66–0.95; p-value 0.01) odds of undergoing stimulation testing.

Analyzing each sex separately, in univariate analysis (Table 3) and multivariate analysis (Table 4) of potential predictors of GH stimulation testing, height Z-score, pubertal status and clinic location were significant predictors for both males and females. The effects of significant predictor variables in the multivariate models are shown in Table 4. Comparison to MPH was a significant predictor only in univariate analysis. Regression models for the sub-population with IGF-I z-scores, controlling for the variables in Table 4, demonstrated a significant effect of IGF-I z-scores for both sexes: females OR 0.50 (95% CI 0.43–0.58), p-value <0.001; males OR 0.44 (95% CI 0.39–0.49), p-value <0.001. Race-ethnicity grouping was a predictor of GH stimulation testing by both univariate and multivariate analysis for females only. Insurance type was a predictor for GH stimulation testing by univariate and multivariate analysis for males only. Unadjusted, males with commercial insurance were less likely to undergo stimulation testing (OR=0.79 [95% CI 0.66–0.94], p-value 0.01). However, adjusted for height z-scores, pubertal status, and clinic location, males with commercial insurance were more likely to undergo stimulation testing than males with government insurance (Table 4).

## Outcomes of GH stimulation testing

Stimulated peak GH levels did not differ by sex (Table 2), with similar proportions of males and females in each peak GH category (Fig. 1). Height z-scores did not vary based on stimulated peak GH category (Fig. 1) or correlate with peak GH concentration, controlling for pubertal status (F-statistic=0.59, p=0.62). The proportion of children with a GH prescription (Fig. 1) or pituitary MRI after stimulation testing did not differ by sex (Table 2).

## GH-prescribed population

Among patients prescribed GH, females were shorter, both by mean height z-score (Fig. 1) and by comparison to MPH z-score (females −1.9 [−2.5, −1.4] vs. males −1.8 [−2.3, −1.3], p-value 0.02); these values were similar to those of the GH stimulation test population. There was no sex difference in IGF-I z-scores among those prescribed GH (females −1.5 [−2.0, −0.5] n=141 vs. males −1.6 [−2.2, −0.8] n=344, N.S.); nor did pubertal status, insurance type, race-ethnicity group, and clinic location differ by sex.

On univariate analysis of the GH stimulation test population (Table 3), stimulated GH level were significant predictors of GH prescription, while height z-score, pubertal status, clinic location, race-ethnicity grouping, and comparison to MPH were not. Controlling for peak stimulated GH level and height z-score, sex was not a predictor of GH prescription (OR 0.89 [95% CI 0.66–1.20], N.S.). In univariate analysis by sex, only peak stimulated GH concentration predicted GH prescription for both females and males (Table 3). Height z-score was a predictor of GH prescription for females, while insurance type predicted GH prescription for males (Table 3). Multivariate analyses with adjusted odds ratio for predictors of GH prescription for each sex are shown in Table 4.

## Discussion

This is the first study to describe sex differences in GH stimulation testing as a factor contributing to sex disparities in pediatric GH treatment [3]. More males than females were evaluated in the endocrine clinic for poor growth/short stature, and a greater proportion of males subsequently underwent GH stimulation testing at heights closer to average than females. However, rates of GH prescribing were similar by sex within peak GH categories reflective of commonly used cutoffs for diagnosis of GHD. GH stimulation test results failed to account for the more severe shortness among tested females, given that peak GH concentrations did not differ by sex or height z-score; this finding likely reflects both the poor specificity of these tests [13] and the myriad non-GHD causes of short stature[21, 22].Yet, sex differences were detected in other clinical and demographic variables that predicted GH stimulation testing and GH prescription.

Our findings in the endocrine clinic population contribute new insights to what is known about sex differences in referral to endocrinology for poor growth/short stature evaluation. The greater number of males seen for subspeciality evaluation in our cohort was comparable to the 2:1 male predominance reported in a 2001 cohort sample [1], indicating little change over time. However, the extent of male predominance was not uniform across racial-ethnic sub-groups. It is known that NHB children have lower odds, and Hispanic and Asian children higher odds, of short stature compared to NHW children in the U.S. [23, 24, 3]. Grimberg *et al* found that a smaller proportion of NHB versus NHW children with growth faltering saw an endocrinologist [2]. Our study further suggests that the rate of evaluation by an endocrinologist may be lower for NHB females when compared with NHB males.

Additionally, in our study population, females not only had shorter stature by height z-score, as was seen in the 2001 study population [1], but also presented in more advanced Tanner stages. The pubertal growth spurt typically occurs between stages 3 and 4 for females, but between stages 4 and 5 for males [25]. Thus, later pubertal stage at presentation leaves less room for intervention for females than for males—this constitutes another potential source of sex disparity in GH treatment. The lack of sex differences in the comparison to MPH, annualized height velocity, and annualized change in height z-score data could suggest some similarity in level of concern at the time of referral between both sexes. However, the large number of missing data points and imprecision of the calculated annualized growth parameter (by virtue of being based on measurements from external providers at varied timepoints) make it hard to draw firm conclusions about this data.

In the GH stimulation test population, a lower rate of testing of females emerged as a factor that may contribute to the previous finding that more males than females were prescribed GH, despite similar rates of short stature between sexes [3]. This held true even when controlling for IGF-I z-scores, which were higher on average among females, in multivariate analysis for the sub-population of referred individuals who had IGF-I z-scores available. GH stimulation testing can contribute to increased GH prescription by two related mechanisms. First, the poor specificity of the stimulation tests, as well as variation in diagnostic threshold used [13], means that more GH stimulation testing leads to more diagnoses of GHD and, thus, more GH prescriptions. The rate of false positive results may be as high 50% [26–29].

Second, because many insurance providers deny coverage for GH treatment of ISS [7], pediatric endocrinologists may order the test to be able to assign a GHD diagnosis to justify GH treatment for children with ISS.

The observed lower odds of undergoing GH stimulation testing for females could reflect sex differences in parent preferences [11, 30], clinician biases [10, 31], and/or insurance related inequities. However, our model controlled for insurance type. Additionally, insurance type predicted likelihood of GH stimulation testing only for males, with commercial insurance associated with higher odds of undergoing stimulation testing. One would expect that if the insurance effect were largely due to commercial insurers being more likely to approve GH stimulation testing and subsequent GH treatment than government insurers, the effect should have been seen also for females. Male predominance found among commercially insured children with GH claims in the U.S. from 2001–2016 supports this interpretation [7]. Taking insurance status as a proxy for socioeconomic status, our finding could reflect greater concern for male stature in affluent families, a notion supported by both qualitative and quantitivative data. A focus-group participant in a study exploring what influences parents to seek evaluation for a child's short stature shared this belief about the impact of affluence [30]. Further, parents wanting their child to undergo pediatric endocrinology evaluation for short stature had higher annual incomes than those of their general community and believed that short stature has negative psychosocial impacts for adult males, but not adult females [32].

The more significant height deficits observed for females than males in the GH stimulation test population could reflect the known preference for tallness in males in the U.S. [33, 11]. This preference could act as an unconscious or conscious factor which lowers the threshold for recommending (in the case of physician) or desiring (in the case of patients and families) GH stimulation testing for males. The height z-scores in the GH-prescribed population were not more sex discrepant than in the GH stimulation test population, suggesting that sex biases related to level of concern about short stature in the endocrinology clinic may have greater effect before GH stimulation testing. This interpretation seems, on the surface, to be supported by the lack of sex differences in GH prescriptions by stimulated peak GH level.

Compared to an analogous study of racial-ethnic differences, this study similarly found sex differences among the endocrine clinic and GH stimulation test populations, but did not find a difference in GH prescriptions, as did the other study [12]. In that study, among children with GH levels in the gray zone of 7–10 ng/ml, the proportion prescribed GH was lower for NHB versus NHW children (53% vs. 77%, p-value 0.04) [12]. While data are lacking to explain this difference, one possible explanation is that differences in provider recommendations, family preferences or a combination of the two are found at the GH prescription stage among racial-ethnic groups, but not between sexes. Nonetheless, we did find sex differences in which variables were associated with GH prescription, which point to more subtle sex differences present at the GH prescription step. Height z-score predicted GH prescription for females, but not males; having commercial insurance was associated with increased odds of having a GH prescription for males, but not females. This latter finding adds to prior observation that children who received GH treatment were more likely to have commercial insurance [3] by suggesting that this effect is more significant for males. The

finding that height z-score did not predict GH prescription for males could be explained by greater importance placed on height for males than females—among both parents [11], especially those of higher socioeconomic status [32, 30], and clinicians [10, 31]—having greater influence than objective measurements of growth. In asking parents of primary care pediatric patients “how short is too short?” responses for male adult heights were on average 5 inches greater than for female adult heights [11]. Further, subjective views of height that have been reported as concerns for parents, such as comparison to peers’ heights or parental expectations [30, 34] and perceptions about their child’s short stature [11, 30, 34], may influence GH prescription more strongly for males, lessening the influence of height z-score in decision-making. That parent and child perception of a child’s height has also been associated with satisfaction of that height lends credence to this hypothesis [35]. Regarding physician practice, in hypothetical scenarios reflective of real-life clinical practice, pediatric endocrinologists were more likely to initiate GH not just for boys compared to girls, but also if the family desired treatment, compared to if the family was neutral, in analyses that took into account growth parameters [10, 36].

Our study findings reinforce that each step in the decision tree of evaluation of short stature potentially can be influenced by sex biases [37]. Critical self-reflection to identify and address sex biases that may impact care is one step clinicians can take to mitigate potential associated harms [38]. Such harms may include under diagnosis of GHD in girls and over treatment of short stature in boys. Examination of how parental, child, and physician attitudes interact and individually contribute to decision-making is needed to shed further light on the observed sex differences in the evaluation and treatment of short stature.

This study was strengthened by the large sample size and the relatively diverse study population. Uniformity of the GH assay and testing protocols avoided variability that would have complicated comparison of GH stimulation testing results. Similarly, height z-scores were based on measurements obtained by a standardized technique, limiting measurement bias. The EHR review method eliminated selection bias that may occur with a participant recruitment approach. However, there were some limitations, mostly related to the retrospective study design. Data on growth velocity, an important aspect of growth evaluations, were not available for every individual and, when available, were based on measurements outside of the endocrinology clinic. Similarly, IGF-I z-score and MPH z-scores were not available for many individuals. Given these limitations, growth velocity, comparison to MPH and IGF-I z-scores could not be included in logistical regression analyses of the total population. Bone ages, another factor in evaluating growth, were not readily extractable from the EHR, nor comparable in a systematic way. Evaluation of the race-ethnicity effect was limited by lower numbers of individuals in the NHB and Hispanic race-ethinicity groupings, as well having to rely on categorizations available in the EHR. More ethnicity categories could have allowed for more nuanced analysis of sex differences by ethnicity. Finally, the generalizability of the study is limited by the single-center study setting, albeit a large sample size derived from a multi-state catchment area. Review of U.S. pediatric GH registry data found differences in the gender ratios by geographic region and GH indication [4], suggestive of regional differences in short stature evaluation.

## Conclusion

Sex differences in pediatric GH treatment seem to reflect sex differences among children seen by pediatric endocrine subspecialists for growth evaluation and sex differences in rates of GH stimulation testing. Rates of GH prescription after stimulation testing did not contribute to sex differences in pediatric GH treatment. However, males seen by the endocrinologists, males who underwent GH stimulation testing and males who were prescribed GH had less severe height deficits than females at each of those steps. Further, height z-score was associated with likelihood of GH prescription for females, but not for males. These results give pause about the extent to which sex biases and preferences related to stature, as opposed to objective growth data, influence medical decision-making in the evaluation and treatment of short stature.

## Acknowledgements

The authors thank Justine Shults and Walter Faig for their assistance with the statistical analyses.

### Statement of Ethics

The study was conducted in compliance with the Declaration of Helsinki. The Institutional Review Board of the Children's Hospital of Philadelphia reviewed the study and determined it met the exemption criteria per 45 CFR 46.104(d) 4(iii). A waiver of HIPAA authorization per 45 CFR 164.512(i)(2)(ii) was granted for accessing identifiable information from the electronic health records (EHR).

### Conflict of Interest Statement

A.G. served as a consultant for the Pediatric Endocrine Society Growth Hormone Deficiency Knowledge Center, sponsored by Sandoz. A.D. served as a consultant for Novo Nordisk, Sandoz, Ascendis, Biomarin and OPKO. A.D. has received research support from Novo Nordisk and Biomarin. A.D. is an Associate Editor of Hormone Research in Paediatrics. The other authors have no conflicts of interest to declare.

### Funding Sources

This study was supported by NIH grant T32 DK063688 from the National Institute of Diabetes and Digestive and Kidney Diseases (C.K.), 2020 Growth Hormone Research Competitive Grant Program Award from Pfizer, Inc. (A.G.), and by NIH grant R01 HD093622 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (A.D.). The funders did not participate in this work.

## Data Availability Statement

The data that support the findings of this study are not publicly available due to their containing information that could compromise the privacy of research participants but are available from C.K. upon reasonable request.

## Abbreviations

| | |
|---|---|
| **CDC** | Centers for Disease Control and Prevention |
| **CI** | confidence interval |
| **EHR** | electronic health record |
| **GH** | growth hormone |
| **GHD** | growth hormone deficiency |

| | |
|---|---|
| **MPH** | sex-adjusted mid-parental height |
| **NHB** | Non-Hispanic Black |
| **NHW** | Non-Hispanic White |
| **OR** | odds ratio |
| **U.S.** | United States of America |

## References


1. Grimberg A, Kutikov JK, Cucchiara AJ. Sex differences in patients referred for evaluation of poor growth. J Pediatr. 2005 Feb;146(2):212–6. [PubMed: 15689911]
2. Grimberg A, Feemster KA, Pati S, Ramos M, Grundmeier R, Cucchiara AJ, et al. Medically underserved girls receive less evaluation for short stature. Pediatrics. 2011 Apr;127(4):696–702. [PubMed: 21422085]
3. Grimberg A, Huerta-Saenz L, Grundmeier R, Ramos MJ, Pati S, Cucchiara AJ, et al. Gender Bias in U.S. Pediatric Growth Hormone Treatment. Sci Rep. 2015 Jun 9;5:11099.
4. Grimberg A, Stewart E, Wajnrajch MP. Gender of pediatric recombinant human growth hormone recipients in the United States and globally. J Clin Endocrinol Metab. 2008 Jun;93(6):2050–6. [PubMed: 18334582]
5. Finkelstein BS, Imperiale TF, Speroff T, Marrero U, Radcliffe DJ, Cuttler L. Effect of growth hormone therapy on height in children with idiopathic short stature: a meta-analysis. Arch Pediatr Adolesc Med. 2002 Mar;156(3):230–40. [PubMed: 11876666]
6. Lee JM, Davis MM, Clark SJ, Hofer TP, Kemper AR. Estimated cost-effectiveness of growth hormone therapy for idiopathic short stature. Arch Pediatr Adolesc Med. 2006 Mar;160(3):263–9. [PubMed: 16520445]
7. Grimberg A, Kanter GP. US Growth Hormone Use in the Idiopathic Short Stature Era: Trends in Insurer Payments and Patient Financial Burden. J Endocr Soc. 2019 Nov 1;3(11):2023–31. [PubMed: 31637343]
8. Allen DB, Backeljauw P, Bidlingmaier M, Biller BM, Boguszewski M, Burman P, et al. GH safety workshop position paper: a critical appraisal of recombinant human GH therapy in children and adults. Eur J Endocrinol. 2016 Feb;174(2):P1–9. [PubMed: 26563978]
9. Brod M, Hojbjerre L, Alolga SL, Beck JF, Wilkinson L, Rasmussen MH. Understanding Treatment Burden for Children Treated for Growth Hormone Deficiency. Patient. 2017 Oct;10(5):653–66. [PubMed: 28386679]
10. Cuttler L, Silvers JB, Singh J, Marrero U, Finkelstein B, Tannin G, et al. Short stature and growth hormone therapy. A national study of physician recommendation patterns. JAMA. 1996 Aug 21;276(7):531–7. [PubMed: 8709401]
11. Cousounis PA, Lipman TH, Ginsburg K, Cucchiara AJ, Grimberg A. How Short is Too Short According to Parents of Primary Care Patients. Endocr Pract. 2014 Nov;20(11):1113–21. [PubMed: 24936551]
12. Hawkes CP, Gunturi H, Dauber A, Hirschhorn JN, Grimberg A. Racial/Ethnic Disparities in the Investigation and Treatment of Growth Hormone Deficiency. J Pediatr. 2021 Apr 23.
13. Kamoun C, Hawkes CP, Grimberg A. Provocative growth hormone testing in children: how did we get here and where do we go now? J Pediatr Endocrinol Metab. 2021 Apr 12.
14. Grimberg A, DiVall SA, Polychronakos C, Allen DB, Cohen LE, Quintos JB, et al. Guidelines for Growth Hormone and Insulin-Like Growth Factor-I Treatment in Children and Adolescents: Growth Hormone Deficiency, Idiopathic Short Stature, and Primary Insulin-Like Growth Factor-I Deficiency. Hormone Research in Paediatrics. 2016;86(6):361–97. [PubMed: 27884013]
15. Grimberg A, Allen DB. Growth hormone treatment for growth hormone deficiency and idiopathic short stature: new guidelines shaped by the presence and absence of evidence. Curr Opin Pediatr. 2017 Aug;29(4):466–71. [PubMed: 28525404]

16. Collett-Solberg P, Ambler G, Backeljauw P, Bidlingmaier M, Biller B, Boguszewski M, 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.
17. Clemmons DR. Consensus statement on the standardization and evaluation of growth hormone and insulin-like growth factor assays. Clin Chem. 2011 Apr;57(4):555–9. [PubMed: 21285256]
18. Center for Disease Control and Prevention. 2000 Center for Disease Control growth charts: United States. http://www.cdc.gov/growthcharts/.
19. Lipman TH, Cousounis P, Grundmeier RW, Massey J, Cucchiara AJ, Stallings VA, et al. Electronic Health Record Mid-Parental Height Auto-Calculator for Growth Assessment in Primary Care. Clin Pediatr (Phila). 2016 Oct;55(12):1100–6. [PubMed: 26507248]
20. The Urban and Rural Classifications. In: Census Bot, editor. Geographic Areas Reference Manual. U.S. Department of Commerce Economics and Statistics Administration 1994.
21. Inzaghi E, Reiter E, Cianfarani S. The Challenge of Defining and Investigating the Causes of Idiopathic Short Stature and Finding an Effective Therapy. Horm Res Paediatr. 2019;92(2):71–83. [PubMed: 31578025]
22. Wit JM, Joustra SD, Losekoot M, van Duyvenvoorde HA, de Bruin C. Differential Diagnosis of the Short IGF-I-Deficient Child with Apparently Normal Growth Hormone Secretion. Horm Res Paediatr. 2021;94(3–4):81–104. [PubMed: 34091447]
23. Komlos J, Breitfelder A. Differences in the physical growth of US-born black and white children and adolescents ages 2–19, born 1942–2002. Ann Hum Biol. 2008 Jan-Feb;35(1):11–21. [PubMed: 18274922]
24. Komlos J, Breitfelder A. Height of US-born non-Hispanic children and adolescents ages 2–19, born 1942–2002 in the NHANES samples. Am J Hum Biol. 2008 Jan-Feb;20(1):66–71. [PubMed: 17941038]
25. Tanner JM. Issues and advances in adolescent growth and development. J Adolesc Health Care. 1987 Nov;8(6):470–8. [PubMed: 3121548]
26. Marin G, Domene H, Barnes K, Blackwell B, Cassorla F, Cutler GJ. 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. 1994;79:537–41. [PubMed: 8045974]
27. Ghigo E, Bellone J, Aimaretti G, Bellone S, Loche S, Cappa M, 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):3323–27. [PubMed: 8784091]
28. Mauras N, Walton P, Nicar M, Welch S, Rogol AD. Growth Hormone Stimulation Testing in Both Short and Normal Statured Children: Use of an Immunofunctional Assay. Pediatric Research. 2000 2000/11/01;48(5):614–18. [PubMed: 11044480]
29. Hilczer M, Smyczynska J, Lewinski A. Limitations of clinical utility of growth hormone stimulating tests in diagnosing children with short stature. Endocrine Regulations. 2006;40(3):69–75. [PubMed: 17100548]
30. Grimberg A, Cousounis P, Cucchiara AJ, Lipman TH, Ginsburg KR. Parental Concerns Influencing Decisions to Seek Medical Care for a Child's Short Stature. Horm Res Paediatr. 2015;84(5):338–48. [PubMed: 26448482]
31. Smuel K, Yeshayahu Y. "Real-world" pediatric endocrine practice; how much is it influenced by physician's gender and region of practice. Results of an international survey. J Eval Clin Pract. 2017 Aug;23(4):866–69. [PubMed: 28585354]
32. Finkelstein BS, Singh J, Silvers JB, Marrero U, Neuhauser D, Cuttler L. Patient attitudes and preferences regarding treatment: GH therapy for childhood short stature. Horm Res. 1999;51 Suppl 1:67–72. [PubMed: 10393494]
33. Hall SS. Size Matters: How Height Affects the Health, Happiness, and Success of Boys-- and the Men They Become. Houghton Mifflin Harcourt; 2006.
34. Hitt T, Ginsburg KR, Cousounis P, Lipman TH, Cucchiara AJ, Stallings VA, et al. Concerns and Expectations of Parents Seeking Subspecialist Care for Their Child's Short Stature. Horm Res Paediatr. 2019;92(5):311–18. [PubMed: 32229729]

35. Hunt L, Hazen RA, Sandberg DE. Perceived versus measured height. Which is the stronger predictor of psychosocial functioning? Horm Res. 2000;53(3):129–38. [PubMed: 11044793]

36. Silvers JB, Marinova D, Mercer MB, Connors A, Cuttler L. A national study of physician recommendations to initiate and discontinue growth hormone for short stature. Pediatrics. 2010 Sep;126(3):468–76. [PubMed: 20805144]

37. Halas JG, Grimberg A. Dilemmas of growth hormone treatment for GH deficiency and idiopathic short stature: defining, distinguishing, and deciding. Minerva Pediatr. 2020 Jun;72(3):206–25. [PubMed: 32274914]

38. Gopal DP, Chetty U, O'Donnell P, Gajria C, Blackadder-Weinstein J. Implicit bias in healthcare: clinical practice, research and decision making. Future Healthc J. 2021 Mar;8(1):40–48. [PubMed: 33791459]

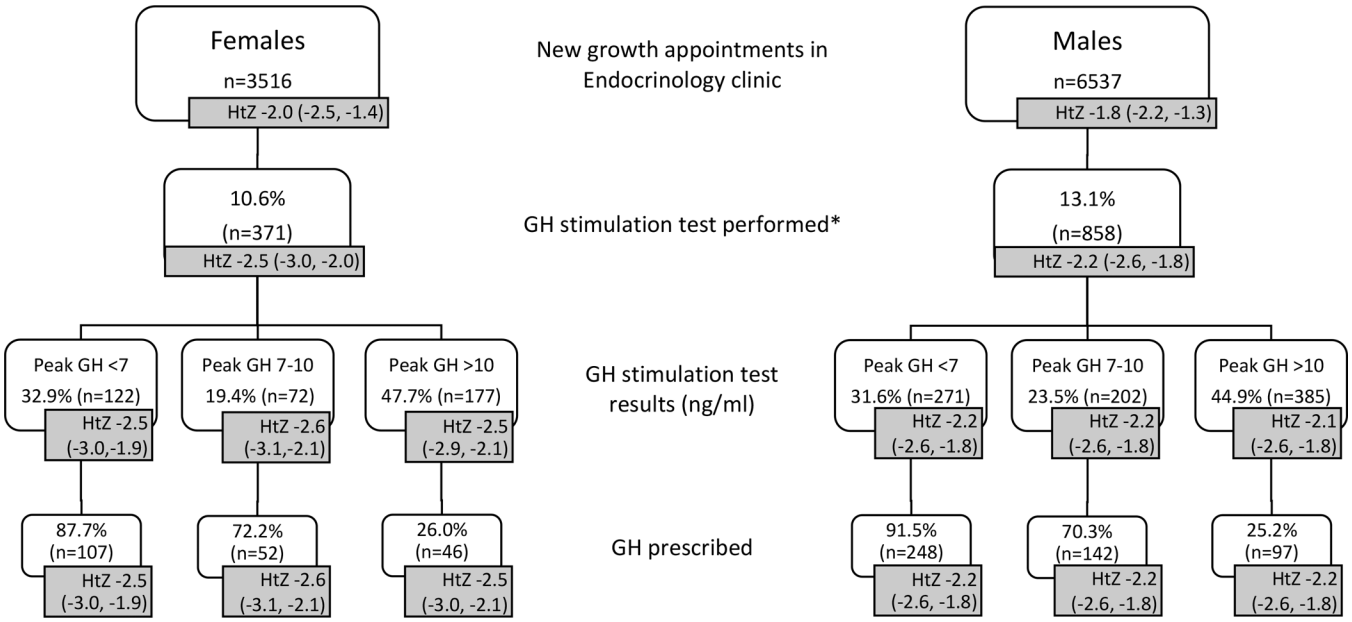


**Fig. 1.**

Sex differences in clinical trajectory and associated height Z-scores. The difference in height z-scores between males and females in all matched sub-groups is statistically significant (p-value <0.001). *Difference between proportions of males and females is statistically significant with p-value <0.001. GH = growth hormone; HtZ = height Z-score, presented as median (interquartile range)

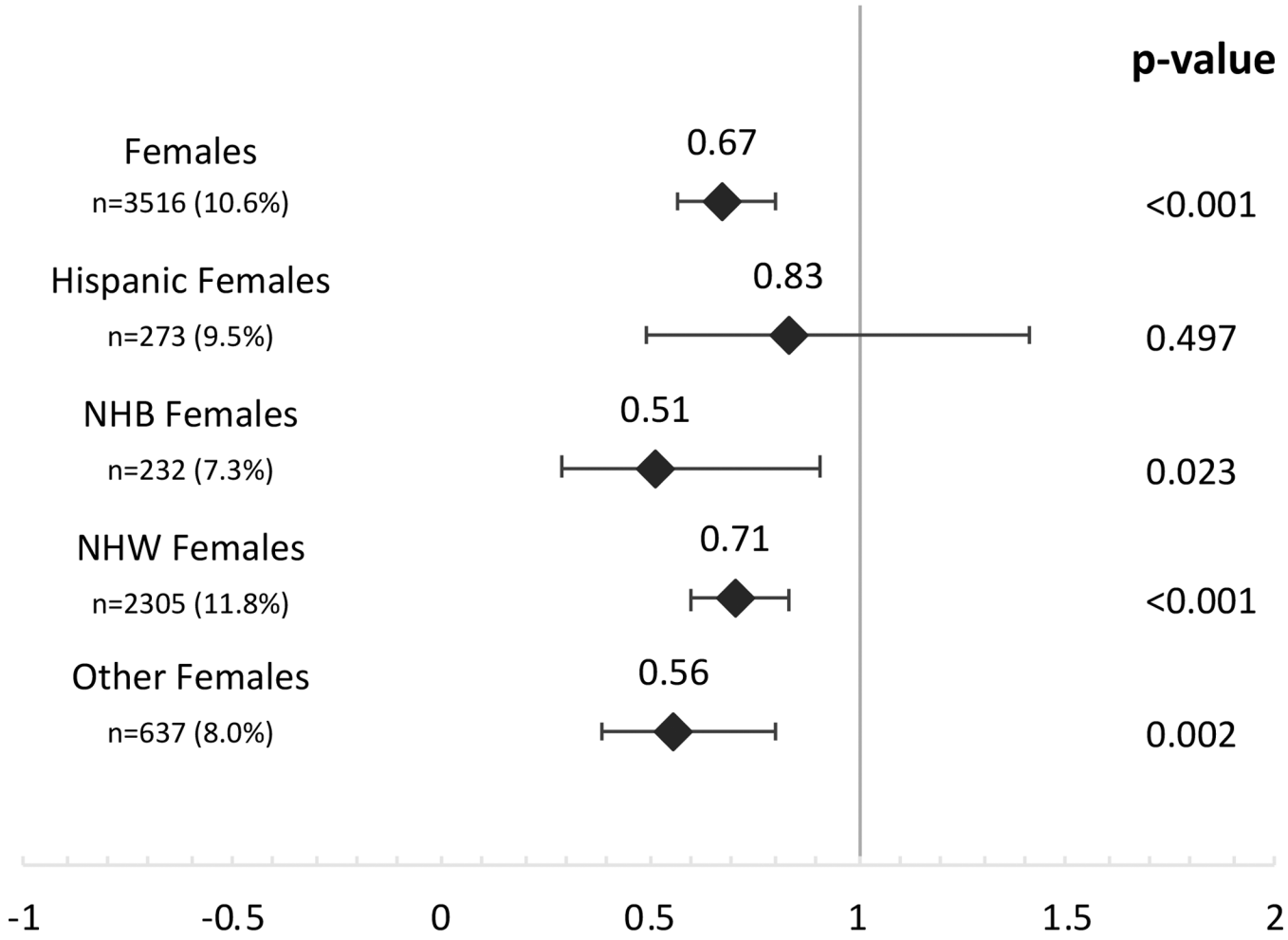


**Fig. 2.**

Odds ratio (OR) of undergoing GH stimulation testing for females compared to males, altogether and by race-ethnicity grouping. Percent of females evaluated in the endocrine clinic who underwent GH stimulation testing is noted under each race-ethnicity grouping label.

**Table 1.**

Demographics and clinical characteristics of endocrine clinic population by sex

| Variable | Females (N=3516) | Males (N=6537) | p-value |
|---|---|---|---|
| **Clinical Outcomes** | | | |
| GH stimulation test performed | 371 (10.6) | 858 (13.1) | **<0.001** |
| **Demographics** | | | |
| Age at clinic visit, years | 10.4 (7.8, 12.5) | 11.8 (8.3, 13.9) | **<0.001** |
| Government insurance | 711 (20.3) | 1142 (17.5) | **<0.001** |
| Urban clinic location | 1284 (36.5) | 2302 (35.2) | 0.19 |
| Race/Ethnicity – *Male:Female* | | | |
| Hispanic – *1.5* | 273 (7.8) | 414 (6.3) | **0.001** * |
| Non-Hispanic Black – *2.4* | 232 (6.6) | 555 (8.5) | **0.005** * |
| Non-Hispanic White – *1.9* | 2305 (65.6) | 4481 (68.5) | 0.83 * |
| Other – *1.6* | 637 (18.1) | 990 (15.1) | **<0.0001** * |
| Unknown | 69 (2.0) | 97 (1.5) | **0.04** * |
| **Growth Parameters** | | | |
| Pubertal stag^ | | | |
| Prepubertal | 1864 (56.4) | 3651 (59.4) | **0.0045** * |
| Early puberty | 991 (30.0) | 2053 (33.4) | **<0.001** * |
| Late puberty | 451 (13.6) | 442 (7.2) | **<0.0001** * |
| Height z-score | (n=3474) −2.0 (−2.5, −1.4) | (n=6462) −1.8 (−2.2, −1.3) | **<0.001** |
| Height z-score minus mid-parental height z-score | (n=2834) −1.5 (−2.2, −0.9) | (n=5347) −1.5 (−2.0, −1.0) | 0.76 |
| Annualized height velocity prior to visit, cm/year | (n=1966) 5.1 (3.9, 6.3) | (n=3777) 4.9 (3.8, 6.2) | **0.02** |
| Annualized change in height z-score prior to visit | (n=1966) −0.1 (−0.3, 0.1) (range −2.3, 4.2) | (n=3777) −0.1 (−0.3, 0.1) (range −2.2, 6.1) | **0.02** |
| Body Mass Index z-score | (n=3453) −0.2 (−1.0, 0.6) | (n=6427) −0.2 (−1.0, 0.5) | 0.05 |
| IGF-I z-score | (n=1570) −0.7 (−1.4, 0.1) | (n=3187) −0.9 (−1.6, −0.1) | **<0.001** |

Continuous variables are presented as median (interquartile range) with p-value for Wilcoxon rank-sum test. Categorical variables are presented as n (%) with p-values for Pearson's Chi-squared test, unless otherwise noted. Small n values represent number of participants with data for corresponding continuous variable if less than N.

* Two-sided t-test of proportions.

^ Prepubertal is defined as Tanner Stage 1, Early puberty as Tanner Stage 1.5 to 3, Late puberty as Tanner Stage 3.5 to 5.

**Table 2.**

Outcomes in the GH stimulation test population

| Variable | Female | Male | p-value |
|---|---|---|---|
| **N** | 371 | 858 | |
| **Peak GH (ng/mL)** | 9.6 (6.0, 13.6) | 9.4 (6.1, 13.2) | 0.67 |
| **GH prescribed** | 205 (55.3) | 487 (56.8) | 0.63 |
| **Pituitary MRI performed** | 148 (39.9) | 332 (38.7) | 0.69 |

Continuous variables presented as median (interquartile range) with p-value for Wilcoxon rank-sum test. Categorical variables presented as n (%) with p-values for Pearson's Chi-squared test.

**Table 3.**

Univariate analysis of predictors of study outcomes

| | Odds Ratio (95% CI) | p-value | Odds Ratio (95% CI) | p-value | Odds Ratio (95% CI) | p-value |
|---|---|---|---|---|---|---|
| **GH stimulation testing** | **All** | | **Females** | | **Males** | |
| **Female** | 0.78 (0.67–0.89) | **<0.001** | --- | | --- | |
| **Height z-score** * | 0.45 (0.42–0.49) | **<0.001** | 0.44 (0.39–0.51) | **<0.001** | 0.42 (0.34–0.46) | **<0.001** |
| **Pubertal stage** | | **<0.001** | | **<0.001** | | **<0.001** |
| Prepubertal (comparator) | 1 | | 1 | | 1 | |
| Early puberty | 0.53 (0.46–0.61) | | 0.57 (0.44–0.74) | | 0.51 (0.43–0.61) | |
| Late puberty | 0.12 (0.08–0.19) | | 0.15 (0.08–0.26) | | 0.10 (0.05–0.20) | |
| **Urban clinic location** | | **<0.001** | | **<0.001** | | **<0.001** |
| Non-urban (comparator) | 1 | | 1 | | 1 | |
| Urban | 1.49 (1.32–1.68) | | 1.62 (1.30–2.01) | | 1.44 (1.24–1.66) | |
| **Race/Ethnicity** | | **0.02** | | **0.01** | | 0.52 |
| NHB (comparator) | 1 | | 1 | | | |
| Hispanic | 0.95 (0.68–1.31) | | 1.33 (0.70–2.52) | | 0.83 (0.61–1.14) | |
| NHW | 1.17 (0.93–1.48) | | 1.70 (1.02–2.83) | | 0.95 (0.73–1.23) | |
| Other | 0.90 (0.69–1.19) | | 1.10 (0.62–1.95) | | 0.85 (0.69–1.05) | |
| **Insurance** | | **0.04** | | 0.81 | | **0.01** |
| Government (comparator) | 1 | | 1 | | 1 | |
| Commercial | 0.85 (0.74–0.99) | | 0.97 (0.74–1.26) | | 0.79 (0.66–0.95) | |
| **Height z-score minus mid-parental height z-score** | 0.58 (0.54–0.62) | **<0.001** | 0.59 (0.53–0.67) | **<0.001** | 0.56 (0.51–0.61) | **<0.001** |
| **IGF-I z-score** * | 0.46 (0.42–0.50) | **<0.001** | 0.50 (0.44–0.58) | **<0.001** | 0.44 (0.40–0.49) | **<0.001** |
| **GH prescription** | **All** | | **Females** | | **Males** | |
| **Female** | 0.94 (0.74–1.20) | 0.63 | --- | | --- | |
| **Peak GH level** * | 0.76 (0.74–0.79) | **<0.001** | 0.77 (0.73–0.82) | **<0.001** | 0.76 (0.73–0.79) | **<0.001** |
| **Height z-score** * | 0.90 (0.77–1.05) | 0.18 | 0.72 (0.55–0.94) | **0.02** | 1.00 (0.82–1.21) | 0.99 |
| **Pubertal stage** | | 0.56 | | 0.76 | | 0.35 |
| Prepubertal (comparator) | 1 | | 1 | | 1 | |
| Early puberty | 0.87 (0.67–1.14) | | 1.09 (0.67–1.79) | | 0.79 (0.57–1.09) | |
| Late puberty | 1.13 (0.46–2.78) | | 1.53 (0.44–5.31) | | 0.86 (0.23–3.21) | |

| | Odds Ratio (95% CI) | p-value | Odds Ratio (95% CI) | p-value | Odds Ratio (95% CI) | p-value |
|---|---|---|---|---|---|---|
| **Urban clinic location** | | 0.05 | | 0.42 | | 0.06 |
| Non-urban (comparator) | 1 | | 1 | | 1 | |
| Urban | 1.26 (1.00–1.58) | | 1.18 (0.79–1.79) | | 1.29 (0.98–1.70) | |
| **Race/Ethnicity** | | 0.11 | | 0.41 | | 0.15 |
| NHB (comparator) | 1 | | 1 | | 1 | |
| Hispanic | 1.39 (0.74–2.59) | | 2.29 (0.66–7.96) | | 1.22 (0.58–2.56) | |
| NHW | 1.24 (0.80–1.91) | | 1.88 (0.69–5.08) | | 1.13 (0.69–1.85) | |
| Other | 0.83 (0.50–1.39) | | 1.27 (0.42–3.86) | | 0.76 (0.42–1.37) | |
| **Insurance** | | 0.07 | | 0.89 | | **0.04** |
| Government (comparator) | 1 | | 1 | | 1 | |
| Commercial | 1.30 (0.98–1.71) | | 1.04 (0.63–1.72) | | 1.43 (1.02–1.99) | |
| **IGF-I z-score** * | 1.05 (0.91–1.20) | 0.51 | 0.96 (0.76–1.22) | 0.75 | 1.10 (0.93–1.29) | 0.28 |
| **Height z-score minus mid-parental height z-score** * | 0.98 (0.85–1.13) | 0.80 | 0.93 (0.74–1.17) | 0.52 | 1.01 (0.85–1.20) | 0.93 |

*Change in odds of GH stimulation testing for every 1-point increase in continuous variable.

CI=confidence interval; NHB=Non-Hispanic Black; NHW=Non-Hispanic White

**Table 4.**

Multivariate logistic regression modeling of predictors of study outcomes by sex

| | Odds Ratio (95% CI) | p-value | Odds Ratio (95% CI) | p-value |
|---|---|---|---|---|
| **GH stimulation testing** | **Females** | | **Males** | |
| **Height z-score** * | 0.46 (0.40–0.53) | **<0.001** | 0.45 (0.41–0.50) | **<0.001** |
| **Pubertal stage** | | | | |
| Prepubertal (comparator) | 1 | | 1 | |
| Early puberty | 0.73 (0.56–0.96) | **0.02** | 0.62 (0.52–0.74) | **<0.001** |
| Late puberty | 0.20 (0.11–0.37) | **<0.001** | 0.16 (0.08–0.31) | **<0.001** |
| **Urban clinic location** | | | | |
| Non-urban (comparator) | 1 | | 1 | |
| Urban | 1.51 (1.20–1.90) | **<0.001** | 1.36 (1.17–1.59) | **<0.001** |
| **Race/Ethnicity** | | | | |
| NHB (comparator) | 1 | | | |
| Hispanic | 1.46 (0.75–2.83) | 0.26 | --- | --- |
| NHW | 1.98 (1.16–3.38) | **0.01** | --- | --- |
| Other | 1.22 (0.67–2.21) | 0.52 | --- | --- |
| **Insurance** | | | | |
| Government (comparator) | --- | --- | 1 | |
| Commercial | --- | --- | 1.26 (1.04–1.54) | **0.02** |
| **GH prescription** | **Females** | | **Males** | |
| **Peak GH level** * | 0.76 (0.71–0.81) | **<0.001** | 0.75 (0.72–0.78) | **<0.001** |
| **Height z-score** * | 0.53 (0.37–0.77) | **0.001** | 0.82 (0.63–1.05) | 0.12 |
| **Insurance** | | | | |
| Government | --- | --- | 1 | |
| Commercial | --- | --- | 2.08 (1.36–3.17) | **0.001** |

Multivariate models were constructed separately for each sex to include variables found to be significant (p-value ≤0.05) in univariate analysis against the outcome of interest and to control for variables clinically significant to the outcomes of interest. P-values refer to statistical significance of OR for each variable in the multivariate models. Overall p-value for all models was <0.001.

* Change in odds of GH stimulation testing for every 1-point increase in continuous variable.

CI=confidence interval; NHB=Non-Hispanic Black; NHW=Non- Hispanic White