



# Psychometric Validation of the Growth Hormone Deficiency-Child Impact Measure (GHD-CIM)

Meryl Brod<sup>1</sup> · Michael Højby Rasmussen<sup>2</sup> · Knud Vad<sup>2</sup> · Suzanne Alolga<sup>1</sup> · Donald M. Bushnell<sup>3</sup> · Jacques Bedoin<sup>2</sup> · Aristides Maniatis<sup>4</sup>

Accepted: 20 December 2020 / Published online: 12 January 2021  
© The Author(s) 2021

## Abstract

**Objective** The aim of this study was to perform psychometric testing of the Growth Hormone Deficiency-Child Impact Measure (GHD-CIM): a patient-reported outcome (PRO) for children with GHD aged 9 to < 13 years and an observer-reported outcome (ObsRO) for parents/guardians of children who are unable to answer for themselves.

**Methods** A non-interventional, multicenter, clinic-based study was conducted in 30 private-practice and large institutional sites in the US and the UK. Psychometric analyses were conducted following an a priori validation statistical analysis plan.

**Results** A preliminary examination of the data determined a PRO version for children aged 9 to < 13 years was not psychometrically sound and therefore the decision was made to have only an ObsRO measure of the GHD-CIM, which would be suitable for children aged 4 to < 13 years. The GHD-CIM ObsRO validity analyses included 98 parents/guardians. Factor analyses identified three domains: Physical Functioning (PHYS), Social Well-Being (SWB), and Emotional Well-Being (EWB). Internal consistency reliability was acceptable for all domains and for the overall score (Cronbach's alpha > 0.70), as was test–retest reliability for the SWB, EWB and overall (above 0.70). At least one convergent validity hypotheses for each domain and overall was proven ( $r > 0.40$ ). Known-groups validity hypotheses for the EWB and SWB domains were significant ( $p < 0.05$ ). Associated effect sizes ranged from  $-0.40$  to  $-0.58$ , indicating that the GHD-CIM is sensitive to change. Anchor-based patient and clinician ratings of severity of disease suggest a preliminary minimally important difference of 5 points for the overall score, and 5 for PHYS, 7 for EWB, and 5 for SWB.

**Conclusions** The GHD-CIM ObsRO was found to be a reliable and valid measure to assess disease-specific functioning, which will provide a more complete patient-centric picture to the growth hormone therapy experience in children.

**Trial Registration** ClinicalTrials.gov NCT02580032, first posted 20 October 2015.

## 1 Background

Growth hormone deficiency (GHD) results when the pituitary gland does not produce enough growth hormone (GH) to stimulate the body to grow, and manifests as an abnormally slow rate of growth in both early and later childhood [1]. Childhood GHD prevalence is within the range of 1.8–2.9 per 10,000 in Europe and the US [2–4]. The

most common cause of childhood-onset GHD is not known [1], but it is classified as idiopathic isolated GHD, without any pituitary structural abnormalities or other concomitant hormone deficiencies. GHD impacts multiple aspects of daily life for children with the condition. The burden of GHD includes symptoms beyond small stature and reduced growth, including poor energy, decreased muscle strength/endurance, other physical limitations, and social and/or emotional impacts [5]. With GH replacement treatment, many children with GHD can reach a normal height for their family [3]. Treatment involves GH injections, typically administered once daily and continuing until the child reaches their final height. Recombinant GH has been in use clinically for over 30 years and is generally considered safe, with rare major adverse effects [6] and only potential minor adverse effects (for example, rash and pain at the injection site) [7].

✉ Meryl Brod  
mbrod@thebrodgroup.net

<sup>1</sup> The Brod Group, Mill Valley, CA, USA

<sup>2</sup> Novo Nordisk A/S, Søborg, Denmark

<sup>3</sup> Health Research Associates, Inc, Mountlake Terrace, WA, USA

<sup>4</sup> Rocky Mountain Pediatric Endocrinology, Centennial, CO, USA

### Key Points for Decision Makers

The Growth Hormone Deficiency-Child Impact Measure (GHD-CIM) is a unique observer-reported outcome (ObsRO) measure developed to assess the impact of growth hormone deficiency (GHD) on children and adolescents in the domains of physical functioning and social and emotional well-being based on the parent/guardians' observations of the child's daily life and health.

The GHD-CIM was found to be a reliable and valid ObsRO measure for parents of children with GHD aged 4 to < 13 years.

Incorporating the brief, 11-item GHD-CIM assessment into clinical practice could allow clinicians to assess a patient's response to therapy.

Unfortunately, no disease-specific measures exist to assess the impact of GHD on children and adolescents. Although the Quality of Life in Short Stature Youth (QoLISSY) questionnaire is also used to assess outcomes in GHD, it has been validated for children with short stature and is not specific to child GHD. To address this gap, the Growth Hormone Deficiency-Child Impact Measure (GHD-CIM) was developed as a disease impact measure. Development of the GHD-CIM included systematic reviews of the GHD literature and existing patient-reported outcome (PRO), observer-reported outcome (ObsRO), and clinician-reported outcome instruments, expert advice, and direct patient and parent/guardian input through qualitative concept elicitation and cognitive debriefing interviews conducted by trained individuals with backgrounds in qualitative interviewing and in the native language of the host country [5].

Based on the conceptual development phase, a 33-item validation-ready measure was developed that was intended to examine the option to have two versions: a PRO for children with GHD aged 9 to < 13 years, and an ObsRO completed by parents/guardians of children with GHD aged 4 to < 9 years [5]. Examination of the validation data would be used to make the determination as to the appropriateness of a PRO for children with GHD aged 9 to < 13 years.

The initial review of the validation study data found that the child data had high ceiling effects not seen in the observer data, other than in the Physical Functioning (PHYS) domain; however, the PHYS domain had poor correlation between observer and child. Considering all of the data in its entirety (ceiling effects, item functioning, inconsistency of findings), it was determined that a PRO version for children aged 9 to < 13 years was not psychometrically sound and therefore the decision was made to have only an

ObsRO measure of the GHD-CIM, which would be suitable for children aged 4 to < 13 years. This manuscript presents the validation data supporting the ObsRO version of the GHD-CIM.

## 2 Methods

The conceptual and psychometric validation of the GHD-CIM followed the scientific principles of PRO and ObsRO measures development according to the US FDA [8] and the European Medicines Agency [9], as well as guidance provided by the International Society for Pharmacoeconomics and Outcomes Research [10]. The study was conducted in accordance with the Declaration of Helsinki [11] and the Guidelines for Good Pharmacoepidemiology Practices [12]. In the US, ethics approval was obtained from the Copernicus Group Institutional Review Board (IRB; tracking number TBG1-15-428). In the UK, independent Research Ethics Committee approval was obtained from the National Health Service Health Research Authority (IRAS Project ID 219425, REC Reference 17/LO/0075). The FDA 21 Code of Federal Regulations §50 and §56 [13] were followed, and signed consent/assent was obtained before any study-related activities were initiated.

### 2.1 Validation Phase

A non-interventional, multicenter, clinic-based study was conducted in 30 private-practice and large institutional (academic/hospital) sites in the US and the UK. The site identified both ongoing and treatment-naïve children with GHD from their current caseload or by chart review, and confirmed their GHD diagnosis and other eligibility criteria (refer to Sect. 2.1.1, for specific information regarding participants). The decision to initiate GH treatment or stay on already initiated treatment was made by the physician as per usual care and independent of the patient's decision to participate in the study. Participants were withdrawn if the child never started treatment, or stopped treatment at any time during the study.

Sites recruited two populations: (1) prepubertal children aged 9 to < 13 years at enrollment with a confirmed GHD diagnosis who answered questions on their own (PRO); and (2) parents/guardians of prepubertal younger children aged 4 to < 9 years at enrollment with a confirmed GHD diagnosis who answered questions as an observer (ObsRO) for children not able to answer for themselves; each population was divided into *treatment-naïve* or *maintenance* groups. Overall, 243 participants (145 children and 98 parents/guardians) were eligible, participated in the validation survey, and were included in the preliminary data analysis used to make the

decision about whether to proceed with a PRO version for children aged 9 to < 13 years.

### 2.1.1 Inclusion and Exclusion Criteria

To be eligible, children in the *treatment-naïve* group had to have (1) a GHD diagnosis previously confirmed by a GH stimulation test, defined as a peak GH level of  $\leq 7.0$  ng/mL; (2) no prior exposure to GH therapy, with a decision made to initiate prescription GH therapy within 3 months; and (3) annualized height velocity below the 25th percentile for chronological age [14]. Children in the *maintenance* group had to have (1) a GHD diagnosis previously confirmed by a GH stimulation test, defined as a peak GH level of  $\leq 10.0$  ng/mL; and (2) be receiving prescription GH therapy for at least 6 months. In addition, all children had to have a body mass index greater than the 5th percentile and less than the 95th percentile.

Parent/guardian participants were eligible if their *treatment-naïve* or *maintenance* child met the diagnostic and medical criteria noted above; however, their child's age had to be 4 to < 9 years. Additionally, all parent/guardian participants were required to live in the same residence as the child with GHD at least 50% of the time.

Child and parent/guardian participants were excluded from the study if they had previously participated in the study, had mental incapacity, unwillingness or language barriers precluding adequate understanding/cooperation, or were likely to be non-compliant in respect to the study conduct, per the physician. Child and parent/guardian participants were also excluded if the child (1) had any clinically significant abnormality likely to affect growth or the ability to evaluate growth, e.g. chromosomal abnormalities and medical syndromes, significant spinal abnormalities; (2) was born small for gestational age (birth weight and/or birth length less than  $-2$  standard deviations (SDs) for gestational age); (3) was diagnosed with diabetes mellitus or fasting blood glucose  $\geq 126$  mg/dL (7.0 mmol/L), or HbA1c  $\geq 6.5\%$  at enrollment; (4) had current inflammatory disease(s) requiring systemic corticosteroid or glucocorticoid treatment for longer than 2 weeks within the last 3 months prior to enrollment; (5) required glucocorticoid therapy and was taking a dose of  $> 400$   $\mu$ g/day of inhaled budesonide or equivalents for longer than 1 month the year prior to enrollment; (6) had concomitant administration of other treatments that may have an effect on growth (hormone replacement therapies were allowed for inclusion); or (7) had any disorder that, per investigator opinion, might jeopardize the subject's safety or protocol compliance.

### 2.1.2 Validation Study Visits and Assessments

At the in-person baseline visit, all participants signed an informed consent/assent and completed a paper validation battery, which included sociodemographic items and relevant medical history (such as age, sex, race, household income, comorbid medical conditions, and GHD history), the GHD-CIM, and the one-item Patient Global Impression of Severity (PGIS) scale. The battery also included the QoLISSY [15], DISABKIDS (DCGM-37) [16], Child Sheehan Disability Scale (CSDS) [17], and Diabetes Fear of Injecting and Self-Testing Questionnaire (D-FISQ) [18]/fear of self-injecting (FSI) subscale. Clinicians completed the Clinician Global Impression of Severity (CGIS) scale and clinical measurements of height/weight.

Follow-up assessments were completed by either a mailed survey sent by clinic staff to *maintenance* participants, including the GHD-CIM and items covering any changes in treatment or major life events since the last assessment, or, for *treatment-naïve* participants, by weekly telephone calls by clinic staff to assess minimal improvement using the Patient Global Rating of Change (PGRC), a two-item scale assessing change and whether change (since treatment start) was meaningful to the child. *Treatment-naïve* participants had up to two additional in-person study visits at their clinic to complete the GHD-CIM, PGIS, and PGRC items covering GHD overall, GHD symptoms, physical functioning, social well-being, and emotions (minimally important difference [MID] assessment and week 12 visits). If an in-person follow-up visit was impossible, clinic staff mailed the week 12 measures to the participant for completion. At follow-up clinic visits, clinicians completed the CGIS and global rating of change, GH treatment start date, and any change in treatment, along with height/weight measurements. Only *treatment-naïve* children who initiated treatment were included in the study.

## 2.2 Statistical Analysis Methods

All analyses were conducted following an a priori validation statistical analysis plan. All statistical tests used a significance level of 0.05 (two-sided) unless otherwise noted. Statistical tests involving multiple comparisons (e.g. analysis of variance [ANOVA] models with multiple groups) included Scheffe post hoc tests. Statistics were conducted using SPSS [19].

### 2.2.1 Sociodemographic and Clinical Characteristics

Descriptive statistics were calculated for demographic and clinical variables to describe the study sample.

### 2.2.2 Descriptive Characteristics: Measure Items

Descriptive statistics were calculated for the individual item responses for items. The ceiling effect threshold for closer examination was set at 50%.

Item-to-item correlation was examined using a Pearson's correlation matrix of each item in the GHD-CIM. A reliability analysis was conducted for all item pairs, and focus was given to correlation coefficients  $> 0.50$ , indicating potential redundancy between the items [20].

Item-to-total correlation was examined using Pearson's correlation between each item score and the total score. A bivariate Pearson's correlation was calculated for each item score against the total score (excluding the item of interest), and any item with a value  $< 0.40$  [21] was examined since this indicates that it may not be sufficiently associated with the remaining items in the hypothesized scale.

### 2.2.3 Rasch Measurement Theory Analysis

Rasch measurement theory analyses were used to examine the ordering of item response options and the scale unidimensionality. Analyzing data according to the Rasch model provides a range of details for checking whether or not summing the scores is justified by the data.

### 2.2.4 Item Reduction

Items were considered for deletion for reasons of high correlations with other items or total score, floor or ceiling effects, poor fit, or conceptual relevance considerations.

### 2.2.5 Factor Analyses

Exploratory factor analysis procedures were performed on the correlation matrices derived from the items comprising the measure. Rotational methods (Varimax with Kaiser Normalization) were employed to achieve a meaningful set of factors. The appropriate number of factors to be extracted was determined as a function of the proportion of common variance accounted for, residuals analysis, and scree plot examination, along with clinical and theoretical interpretability. Standardized factor loadings of at least 0.40 were considered acceptable.

A confirmatory factor analysis was also conducted to verify the final factor structure. The following fit indices were used to test and confirm the relationship between the observed variables and their underlying latent constructs: comparative fit index, goodness-of-fit index, and root mean square error of approximation.

### 2.2.6 Reliability

Cronbach's alpha was used to assess internal consistency reliability [22]. This statistic is used to analyze additive scales to determine to what degree the items within the scale are associated. A high internal consistency suggests that the scale or subdomain is measuring a single construct. Alpha values range from 0.00 to 1.00; however, a minimum correlation of 0.70 is necessary to claim the instrument is internally consistent (alpha values between 0.80 and 0.90 are preferred).

Test-retest reliability was administered approximately 2 weeks after baseline and assessed using the intraclass correlation coefficient (ICC; 2-way mixed-effect model with absolute agreement) in a subsample from the *maintenance* group who indicated experiencing no change on the Changes Since Last Assessment items (major life events or treatment).

### 2.2.7 Convergent Validity

To assess convergent construct validity for each domain and total score of the measure, Spearman correlations (due to non-normal distributions) were computed to measure the association between GHD-CIM scores and the other measures included in the study. Convergent validity was considered supported when scores were substantially correlated ( $\geq 0.40$ ) with items or instruments measuring similar concepts. When more than one hypothesis per domain was proposed, at minimum one hypothesis should be met to claim convergent validity has been shown. The a priori hypotheses that were tested were:

- GHD-CIM Total will be correlated with the QoLISSY Total score.
- GHD-CIM Total will be correlated with overall GHD interference rating.
- GHD-CIM Physical Functioning (PHYS) will be correlated with the QoLISSY Physical score.
- GHD-CIM PHYS will be correlated with the overall physical functioning rating.
- GHD-CIM Social Well-Being (SWB) will be correlated with the QoLISSY Social score.
- GHD-CIM SWB will be correlated with the overall social well-being rating.
- GHD-CIM Emotional Well-being (EWB) will be correlated with the DISABKIDS Emotional score.
- GHD-CIM EWB will be correlated with the overall emotional well-being rating.

### 2.2.8 Known-Groups Validity

Known-groups validity was also tested for each domain and the total score based on a priori hypotheses using a two-tailed test at a  $p < 0.05$  level. The a priori hypotheses that were tested were:

- Total: Children (*maintenance* group) who start GH treatment earlier will have better total scores on the GHD-CIM.
- PHYS: Increases in height (*treatment-naïve*) are significantly related to greater improvements in physical functioning.
- SWB: Children with better coping related to their height will have better social well-being.
- EWB: Children with more positive emotions related to their height will have greater emotional well-being.

### 2.2.9 Sensitivity to Change

Formal responsiveness was not assessed within this study's protocol as it was neither a treatment intervention nor a randomized clinical trial. However, to assess potential sensitivity to changes, we examined the 12-week GHD-CIM follow-up scores of the *treatment-naïve* group who initiated treatment based on usual care. This provides an idea of the magnitude of change from baseline to follow-up on a new GH treatment. An exploratory analysis of sensitivity to change was conducted using distributional methods to evaluate the effect size (mean change score divided by the SD of the baseline score). Higher values for the effect size indicated a greater sensitivity to change. Using the preferred approach [23], standardized effect size, the mean change divided by an SD served as an effect size index. Standardized effect sizes of 0.2–0.5 were regarded as 'small', 0.5–0.8 were regarded as 'moderate' and those above 0.8 were regarded as 'large' [24].

### 2.2.10 Interpretation of Meaningful Change

To examine meaningful within-patient change, anchor-based methods were used, with the primary anchor being subjective perceptions of disease severity (PGIS), but also examining more objective, clinician perceptions (CGIS). This analysis used only the *treatment-naïve* patients who indicated having an improvement in these anchors. Meaningful change was defined as the difference between these two momentary assessments of GHD severity, with differences anchored to changes in one response option (e.g. severe to moderate) or two response options (e.g. severe to mild) [25].

### 2.2.11 Scoring

Factor analysis was conducted, which informs as to the measurement model and domain structure. The GHD-CIM is scored by summing the items for each domain and converting to a 0- to 100-point scale, with higher scores representing a greater impact. Three positively framed items (PHYS) were reverse-scored. Missing items ('Don't Know' responses were treated as missing) are allowed and are accounted for in the scoring, with three of four items needed to score the PHYS, three of four needed for EWB, and two of three needed for the SWB. If there are missing items and the number of missing items does not exceed the number needed to score any domain, then the transformation calculation must be adjusted for the number of items included in the domain score. The overall score is calculated as the mean of the three domain scores (if a domain score could not be calculated due to missing data, then an overall score was not calculated).

## 3 Results

Given the decision was made, after examining the preliminary data, to only proceed with the ObsRO version, a total of 98 parents/guardians were included in the final psychometric validation analysis set used for the findings reported in this study.

### 3.1 Statistical Analysis

#### 3.1.1 Sociodemographic and Clinical Characteristics

Most parent participants were from the US (90.8%). The mean child age was 6.7 years, ranging between 4 and 9 years. The parents' mean age was 38.6 years (range 25–53). Children were predominantly male (65.3%) and White (82.7%).

The mean age of the children at diagnosis was 5.1 years (range 0.1–8), and the mean age when a child first started taking GHD medication was 5.2 years (range 0.1–8). A small percentage of children were taken off GHD medication (4.2%) for an average of 4.3 weeks. Most subjects (79.6%) used a pen for medication injection, with approximately 4.1% using a needle/syringe.

Over half (53.1%) of the children had no other health conditions and less than one-quarter of the children (20.4%) had been prescribed other medications. Table 1 presents the demographic and clinical characteristics of parent participants and their children.



### 3.1.2 Item Reduction

The GHD-CIM was examined for item characteristics, including floor and ceiling effects, missing data, item-to-item correlations, and item-to-total correlations; 22 items were dropped during item reduction. Four of the dropped items, which had conceptual coverage with other items, were positively framed, and, when examining the response patterns, were different than the other items and may have added confusion for children completing the questionnaire. The other 18 dropped items were deemed to be conceptually redundant with other items and/or had high ceiling effects. The final GHD-CIM included 11 items.

### 3.1.3 Descriptive Characteristics of Growth Hormone Deficiency-Child Impact Measure (GHD-CIM) Items

The full range of response options (0–4) were used for 8 of the 11 items. For the items ‘Energy’ and ‘Upset’, only responses 0–2 were used, as no-one answered ‘None/A Little’ for Energy or ‘All of the Time/Often’ for Upset, and for the item ‘Treated Differently by Children’ only responses 0–3 were used, as no-one answered ‘All of the Time’. Despite using the full range of responses, the overall trend was toward the ‘better’ end of the scale. Ceiling effects, responses of ‘Not at All/Never/None’ (where respondents could not get any better) were evident. Consequently, the means and medians were lower than expected.

Mean scores ranged from 0.50 (how often did your child feel upset?) to 2.15 (how often did people think your child was younger than they are?).

For the total sample, missing data (including missing and ‘Don’t Know’ responses) were minimal.

For most items, item-to-total correlations showed acceptable associations between each item against the remainder of the items as a total score (excluding that item). Two items with lower-than-expected associations ( $< 0.30$ ) were ‘Often Upset’ (0.281) and ‘Worried About Growing’ (0.184).

Table 2 shows the descriptive statistics for individual responses on the GHD-CIM items.

### 3.1.4 Rasch Measurement Theory Analysis

Item thresholds show that most items (30 of 33) were disordered, i.e. the threshold values between adjacent pairs of response options were disordered by magnitude. The person–item distribution showed that while the items covered a wide range (from difficult to not difficult), the persons were more clustered to the right side, indicating a fairly ‘healthy’ population.

### 3.1.5 Factor Analyses

An exploratory factor analysis (principal components analysis) was performed ( $n = 98$ ) on the final 11-item measure. As seen in Table 3, three factors were presented. When evaluating the items within each factor, there was concordance, along with some differences, with the original conceptual framework. All items comprising the EWB and SWB were factored into those domains; however, the items within ‘Symptoms’ and ‘Physical Functioning’ were factored into a single component. It was determined that this was conceptually consistent and that symptom items (e.g. tiredness, energy) could be conceptualized as physical functioning items. This resulted in a change to the hypothesized conceptual framework by combining the ‘Symptoms’ and ‘Physical Functioning’ items into the single PHYS domain.

A post hoc confirmatory factor analysis was also performed on the GHD-CIM using IBM® SPSS® Amos™ [26, 27]. Adequate fit indices were seen, i.e. comparative fit index (0.984), goodness-of-fit index (0.984), root mean square residual (0.0486), and root mean square error of approximation (0.045) [28, 29]. Additionally, a higher order factor analysis was conducted on the three subscale scores to determine the ability to create an overall score of treatment burden. The subscales factored into a single component, with 61.3% of total variance explained.

### 3.1.6 Reliability

All coefficients exceeded the threshold of 0.70, indicating internally consistent scales (see Table 4). Test–retest reliability was assessed using the ICC in a subsample from the *maintenance* group who indicated experiencing no change on the Changes Since Last Assessment items. ICCs for the EWB, SWB, and overall were above 0.70, and were lower than desired for the PHYS (0.66).

### 3.1.7 Convergent Validity

For the eight convergent validity hypotheses, associations were significant for seven of eight comparisons (with PHYS being lower than expected). Six associations were greater than the threshold of 0.40, as expected. Significant correlations over 0.40 were found for the overall score with QoLISSY total score ( $p = -0.78$ ) and overall GHD interference rating (0.44); SWB score with QoLISSY social score ( $p = -0.79$ ) and overall GHD social well-being rating (0.49); and EWB score with DISABKIDS emotional score (0.62) and overall GHD emotional well-being score (0.55).

**Table 1** Demographic and clinical characteristics

	Parent—treatment-naïve [ <i>n</i> = 25]	Parent—maintenance [ <i>n</i> = 73]	Total [ <i>n</i> = 98]
Age of the child, years [mean (range)]	6.5 (4–8)	6.8 (4–9)	6.7 (4–9)
Age of the child, years			
4–8	25 (100.0)	72 (98.6)	97 (99.0)
9–13	0 (0.0)	1 (1.4)	1 (1.0)
Sex of the child			
Female	9 (36.0)	25 (34.2)	34 (34.7)
Male	16 (64.0)	48 (65.8)	64 (65.3)
Child ethnicity <sup>a</sup>			
White	18 (72.0)	63 (86.3)	81 (82.7)
Other <sup>b</sup>	7 (28.0)	10 (13.7)	17 (17.3)
More than one ethnicity	5 (20.0)	7 (9.6)	12 (12.2)
Relationship to the child			
Mother	19 (76.0)	60 (82.2)	79 (80.6)
Father	6 (24.0)	13 (17.8)	19 (19.4)
Parent age, years [mean (range)]	40.5 (29–51)	37.9 (25–53) <sup>c</sup>	38.6 (25–53) <sup>d</sup>
Parent work status			
Work full-time for pay	11 (44.0)	37 (50.7)	48 (49.0)
Work part-time for pay	8 (32.0)	12 (16.4)	20 (20.4)
Student	0 (0.0)	1 (1.4)	1 (1.0)
Student and working	0 (0.0)	3 (4.1)	3 (3.1)
Not working (retired)	0 (0.0)	0 (0.0)	0 (0.0)
Not working (disabled)	0 (0.0)	0 (0.0)	0 (0.0)
Not working (other)	6 (24.0)	20 (27.4)	26 (26.5)
Country			
US	20 (80.0)	69 (94.5)	89 (90.8)
UK	5 (20.0)	4 (5.5)	9 (9.2)
US household income			
Less than \$20,000	0 (0.0) <sup>e</sup>	2 (2.9) <sup>f</sup>	2 (2.2) <sup>g</sup>
\$20,001–\$40,000	0 (0.0) <sup>e</sup>	6 (8.7) <sup>f</sup>	6 (6.7) <sup>g</sup>
\$40,001–\$60,000	0 (0.0) <sup>e</sup>	1 (1.4) <sup>f</sup>	1 (1.1) <sup>g</sup>
\$60,001–\$80,000	2 (10.0) <sup>e</sup>	7 (10.1) <sup>f</sup>	9 (10.1) <sup>g</sup>
\$80,001–\$100,000	2 (10.0) <sup>e</sup>	7 (10.1) <sup>f</sup>	9 (10.1) <sup>g</sup>
More than \$100,000	11 (55.0) <sup>e</sup>	39 (56.5) <sup>f</sup>	50 (56.2) <sup>g</sup>
Declined to answer	5 (25.0) <sup>e</sup>	7 (10.1) <sup>f</sup>	12 (13.5) <sup>g</sup>
UK household income			
Less than £16,000	1 (20.0) <sup>h</sup>	0 (0.0) <sup>i</sup>	1 (11.1) <sup>j</sup>
£16,001–£32,000	1 (20.0) <sup>h</sup>	1 (25.0) <sup>i</sup>	2 (22.2) <sup>j</sup>
£32,001–£48,000	0 (0.0) <sup>h</sup>	2 (50.0) <sup>i</sup>	2 (22.2) <sup>j</sup>
£48,001–£64,000	1 (20.0) <sup>h</sup>	1 (25.0) <sup>i</sup>	2 (22.2) <sup>j</sup>
£64,001–£80,000	0 (0.0) <sup>h</sup>	0 (0.0) <sup>i</sup>	0 (0.0) <sup>j</sup>
More than £80,000	1 (20.0) <sup>h</sup>	0 (0.0) <sup>i</sup>	1 (11.1) <sup>j</sup>
Declined to answer	1 (20.0) <sup>h</sup>	0 (0.0) <sup>i</sup>	1 (11.1) <sup>j</sup>
Child age at diagnosis, years [mean (range)]	6.3 (3–8)	4.6 (0.1–8)	5.1 (0.1–8)
Child age when they first started taking GHD medication, years [mean (range)]	6.5 (4–8)	4.8 (0.1–8)	5.2 (0.1–8)
Child ever taken off GHD medication, yes	NA	3 (4.2) <sup>k</sup>	3 (4.2) <sup>l</sup>
Length of time stopped taking GHD medication, weeks [mean (range)]	NA	4.3 (4–5)	4.3 (4–5)
Medication injection device			
Needle and syringe	0 (0.0)	4 (5.5)	4 (4.1)
Pen	20 (80.0)	58 (79.5)	78 (79.6)

**Table 1** (continued)

	Parent—treatment-naïve [ <i>n</i> = 25]	Parent—maintenance [ <i>n</i> = 73]	Total [ <i>n</i> = 98]
Needle-free injector	0 (0.0)	1 (1.4)	1 (1.0)
Other	5 (20.0)	10 (13.7)	15 (15.3)
Child taking other prescription medications, yes	6 (24.0)	14 (19.2)	20 (20.4)
Child has other health conditions <sup>a</sup>			
Ear, nose, and throat conditions	2 (8.0)	11 (15.5) <sup>k</sup>	13 (13.5) <sup>l</sup>
Endocrine disorders (including thyroid disorders)	4 (16.0)	7 (9.9) <sup>k</sup>	11 (11.5) <sup>l</sup>
Eye disorders	2 (8.0)	3 (4.2) <sup>k</sup>	5 (5.2) <sup>l</sup>
Intellectual, learning, and/or physical disability	1 (4.0)	5 (7.0) <sup>k</sup>	6 (6.3) <sup>l</sup>
Lung disease, respiratory conditions (including allergies and asthma)	4 (16.0)	4 (5.6) <sup>k</sup>	8 (8.3) <sup>l</sup>
Mental health conditions (including ADHD, depression, and anxiety)	1 (4.0)	9 (12.7) <sup>k</sup>	10 (10.4) <sup>l</sup>
Speech and/or language disorders	2 (8.0)	7 (9.9) <sup>k</sup>	9 (9.4) <sup>l</sup>
Other conditions	5 (20.0)	15 (21.1) <sup>k</sup>	20 (20.8) <sup>l</sup>
None	14 (56.0)	37 (52.1) <sup>k</sup>	51 (53.1) <sup>l</sup>

Data are expressed as *n* (%) unless otherwise specified

ADHD attention-deficit hyperactivity disorder, GHD growth hormone deficiency

<sup>a</sup>Participants could choose more than one response

<sup>b</sup>Other ethnicities include Black/African American, Asian/Asian British, Latino or Hispanic, American Indian/Alaskan Native, and Native Hawaiian/Other Pacific Islander, or another group not listed

<sup>c</sup>*N* = 65 since eight participants declined to answer this question

<sup>d</sup>*N* = 90 since eight participants declined to answer this question

<sup>e</sup>*N* = 20 for all US participants

<sup>f</sup>*N* = 69 for all US participants

<sup>g</sup>*N* = 89 for all US participants

<sup>h</sup>*N* = 5 for all UK participants

<sup>i</sup>*N* = 4 for all UK participants

<sup>j</sup>*N* = 9 for all UK participants

<sup>k</sup>*N* = 71 since two participants declined to answer this question

<sup>l</sup>*N* = 96 since two participants declined to answer this question

### 3.1.8 Known-Groups Validity

The known-groups a priori validity hypotheses for SWB and EWB were significant ( $p < 0.05$ ). Additionally, all domains and the total score were able to discriminate between the levels of coping (Fig. 1) and emotional well-being (Fig. 2). For hypotheses of age at treatment initiation, trends (GHD impacts worsening as age at the start of treatment increased) were seen but were non-significant. Hypotheses for improvements in height being associated with better physical functioning were found.

### 3.1.9 Sensitivity to Change

For the change over time for the *treatment-naïve* participants who completed a follow-up assessment 12 weeks post-baseline, marked improvements were noted for the SWB, EWB, and overall (ranging between  $-6.4$  and  $-8.6$  points on a 0- to 100-point scale). The PHYS domain did not show

an improvement over the 12 weeks. Associated effect sizes (mean change divided by the baseline SD) ranged from  $-0.26$  (PHYS) to  $-0.45$  (EWB), indicating that the GHD-CIM is sensitive to change at a moderate level.

### 3.1.10 Interpretation of Meaningful Change

Given the study design (observational with an understanding that changes over time may not be assessed), there was a small sample that indicated some improvement over the 12 weeks. Using the anchors of PGIS and CGIS, GHD-CIM scores were calculated for the groups who had 1- and 2-point improvements in the PGIS and CGIS scales. Table 5 shows that changes for the GHD-CIM total and domain scores, in all but one case (CGIS for SWB), were larger, as expected, for the two-category improvements than for the one-category improvements. Smaller amounts of change, overall, were seen for PHYS. Overall score differences ranged from 5.1 to 8.3 points; PHYS (1.6–7.6 points); EWB (9.5–12.3 points);



**Table 2** GHD-CIM item characteristics (total sample  $n = 98$ )

Item content	<i>N</i>	Mean (SD)	Median	Range	<i>N</i> (%) floor	<i>N</i> (%) ceiling	<i>N</i> (%) missing
In the past week							
1. How physically strong was your child when doing sports or other physical activities	97	1.19 (1.03)	1.00	0–4	2 (2.0)	27 (27.6)	1 (1.0)
2. How tired was your child during the day	98	0.90 (0.91)	1.00	0–4	1 (1.0)	37 (37.8)	0 (0.0)
3. How active was your child	98	0.93 (0.72)	1.00	0–4	1 (1.0)	25 (25.5)	0 (0.0)
4. How much energy did your child have	98	0.93 (0.46)	1.00	0–2	0 (0.0)	14 (14.3)	0 (0.0)
In the past week, <i>because of their size</i> , how often							
5. Did people think your child was younger than they are	93	2.15 (1.53)	2.00	0–4	24 (24.5)	23 (23.5)	5 (5.1)
6. Were they teased	90	0.60 (0.99)	0.00	0–4	2 (2.0)	61 (62.2)	8 (8.2)
7. Were they treated differently by <i>children</i>	93	0.75 (1.01)	0.00	0–3	0 (0.0)	56 (57.1)	5 (5.1)
In the past week, <i>because of their size</i> , how often did your child feel							
8. Embarrassed	92	0.58 (0.86)	0.00	0–4	1 (1.0)	58 (59.2)	6 (6.1)
9. Nervous or anxious	93	0.62 (0.98)	0.00	0–4	1 (1.0)	61 (62.2)	5 (5.1)
10. Worried about growing	92	0.71 (1.03)	0.00	0–4	2 (2.0)	56 (57.1)	6 (6.1)
11. Upset	92	0.50 (0.78)	0.00	0–2	0 (0.0)	62 (63.3)	6 (6.1)

*GHD-CIM* Growth Hormone Deficiency – Child Impact Measure, *SD* standard deviation

**Table 3** GHD-CIM factor analysis<sup>a</sup>

		Factor		
		1	2	3
Emotional well-being	CIM29: size—often upset	0.964	0.042	– 0.077
	CIM19: size—often embarrassed	0.922	0.185	– 0.003
	CIM28: size—often worried about growing	0.907	0.133	– 0.013
	CIM22: size—often nervous/anxious	0.851	0.189	– 0.166
Social well-being	CIM15: size—often teased	0.074	0.863	– 0.160
	CIM12: size—people think child is younger than they are	0.129	0.812	0.108
	CIM17: size—often treated differently by children	0.242	0.810	– 0.161
Physical functioning	CIM10: how active was your child (reversed)	– 0.056	– 0.058	0.756
	CIM08r: how much energy did child have (reversed)	0.011	0.110	0.729
	CIM06: how physically strong was your child (reversed)	– 0.055	– 0.053	0.682
	CIM07: how tired was your child during day	0.090	0.204	– 0.560

*GHD-CIM* Growth Hormone Deficiency – Child Impact Measure

<sup>a</sup>Extraction method: principal component analysis, explaining 69.2% total variation. Rotation method: varimax with Kaiser Normalization

and SWB (4.1–9.2 points). A preliminary estimate of the MID for the overall score is suggested to be 5 points. Similarly, based on these results, the preliminary estimate would be 5 for PHYS, 7 for EWB, and 5 for SWB.

### 3.2 Theoretical Model

Based on the validation study findings, the preliminary theoretical model of the impact of GHD on children was revised and is shown in Fig. 3. The final 11-item GHD-CIM is shown in Fig. 4.

The final validated 11-item GHD-CIM is an ObsRO for parents of children aged 4 to < 13 years.

## 4 Discussion

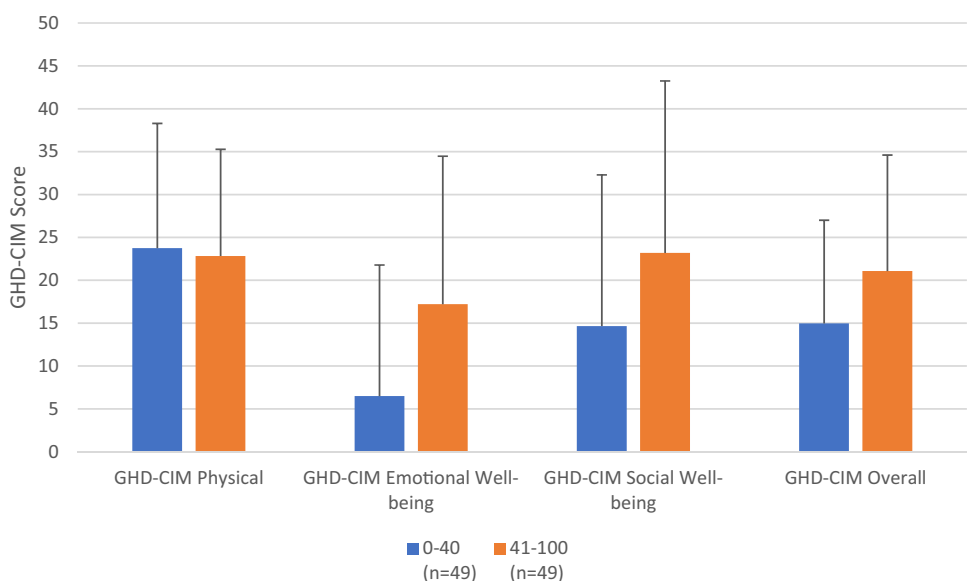
Psychometric analyses for evaluating measurement properties is an iterative process considering both conceptual relevance and psychometric properties. The analyses used to evaluate item performance of the GHD-CIM were in accordance with classical psychometric theory [30].

**Table 4** Evidence for internal consistency of the GHD-CIM

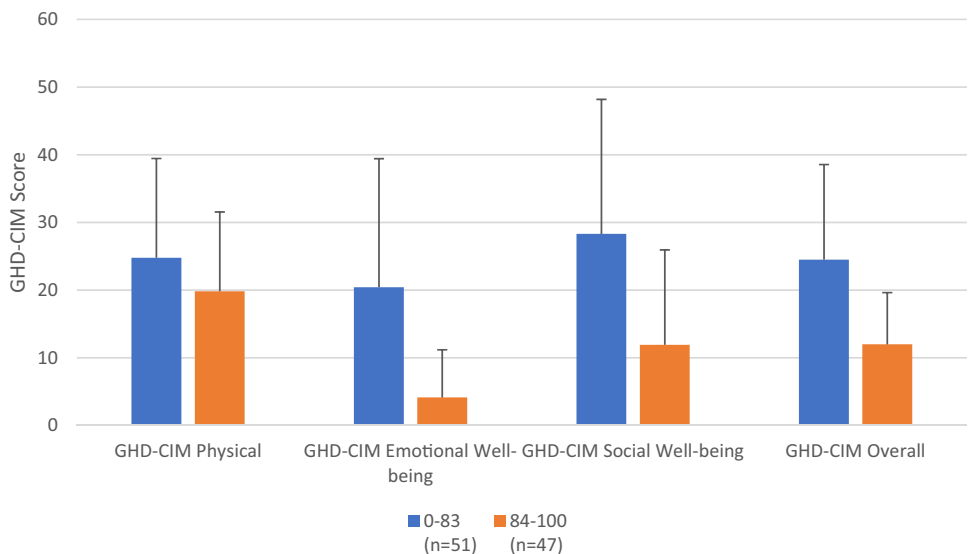
	Cronbach's alpha by domain	Cronbach's alpha of overall score
Internal consistency		
Physical functioning	0.813	0.810
Emotional well-being	0.850	
Social well-being	0.797	
Reproducibility		
Physical functioning ( <i>n</i> = 65)	ICC [95% CI]	
Emotional well-being ( <i>n</i> = 61)	0.662 [0.445–0.794]	
Social well-being ( <i>n</i> = 62)	0.786 [0.667–0.866]	
Overall score ( <i>n</i> = 65)	0.751 [0.618–0.842]	
	0.789 [0.676–0.866]	

*GHD-CIM* Growth Hormone Deficiency-Child Impact Measure, *ICC* intraclass correlation coefficient, *CI* confidence interval

**Fig. 1** Evidence for known-groups validity of the GHD-CIM based on the QoLISSY Coping domain. Significance for emotional well-being and social well-being was *p* < 0.05; physical functioning and overall were not significant. Assessed using the QoLISSY Coping domain. *GHD-CIM* Growth Hormone Deficiency-Child Impact Measure, *QoLISSY* Quality of Life in Short Stature Youth



**Fig. 2** Evidence for known-groups validity of the GHD-CIM based on the QoLISSY Emotional domain. Significance for all GHD-CIM domains was *p* < 0.01. Assessed using the QoLISSY Emotional domain. *GHD-CIM* Growth Hormone Deficiency-Child Impact Measure, *QoLISSY* Quality of Life in Short Stature Youth



**Table 5** Meaningful change thresholds (within-person)

	Physical functioning	Emotional well-being	Social well-being	Overall
PGIS—improve 1 category ( $n = 12$ )	− 1.6 (9.6)	− 9.5 (10.8)	− 4.1 (5.1)	− 5.1 (5.2)
PGIS—improve 2 categories ( $n = 16$ )	− 2.0 (10.1)	− 12.3 (11.7)	− 7.2 (7.7)	− 7.1 (6.4)
CGIS—improve 1 category ( $n = 6$ )	− 5.2 (12.6)	− 9.5 (12.7)	− 9.2 (7.4)	− 8.0 (8.7)
CGIS—improve 2 categories ( $n = 9$ )	− 7.6 (11.1)	− 9.9 (13.7)	− 7.2 (6.7)	− 8.3 (7.7)

CGIS Clinician Global Impression of Severity, PGIS Patient Global Impression of Severity

After examination of the preliminary data, we concluded that given the high ceiling effects for the PRO version, the ObsRO version was more appropriate as a valid and reliable measure of the impact of GHD on children aged 4 to < 13 years. We believe the ceiling effects were indicative of children with GHD having short stature their entire lives and possibly having reduced insight due to accommodation to their condition. Additionally, given that many clinical trials for GHD continue for more than 1 year until a child reaches puberty around age 13 years, a child entering the study at age <9 years, where an ObsRO would be required, will of course be older for future assessments. Using the same ObsRO version over the life of the study for all assessments will facilitate rater consistency over time.

Results showed that the GHD-CIM ObsRO version has acceptable measurement properties of item-to-item, item-to-total correlations, and test–retest reliability; the factor structure was confirmed for a three-domain measure (PHYS, SWB, and EWB), as well as justification for an overall score. Reliability was acceptable, with good internal consistency and adequate to good test–retest reliability. A priori criteria for convergent validity was met for seven of eight domains and the total score (with PHYS being lower than expected). Known-groups validity was confirmed for EWB and SWB, and trends were found for the PHYS and overall score. Emotional functioning was also found to discriminate between all domains and the overall score, with better emotional functioning evidenced for those with less GHD impact.

As with all studies there are limitations that should be noted. First, it was difficult to recruit parents/guardians of children who started treatment at an early age due to US real-world clinical practice patterns. Future studies in countries where treatment is generally started at an earlier age are warranted. Furthermore, this population appears to be on the healthier end of the spectrum, as indicated by the larger-than-expected floor/ceiling effects. With the recruitment of a more diverse population (including more severely symptomatic subjects), the person distribution within the Rasch analysis would be expanded. Additionally, although there was evidence of responsiveness of the GHD-CIM, this was a short study (12 weeks) and treatment benefit may take longer

to manifest. In addition, further analyses with more robust sample sizes within the improvement categories would need to be conducted to establish meaningful change thresholds. In future studies that will be testing change, given the effect size we have seen for the total score ( $-0.44$ ), calculations would dictate a sample size of approximately 122 ( $\alpha = 0.05$ , statistical power at 80%, using percentiles of the effect size distribution). Lastly, it would be helpful to conduct qualitative studies to assess patient perception of meaningful changes and where they occur on the 0–100 scale in studies where change in disease condition is evaluated.

The GHD-CIM is the first GHD-specific measure of the impact of this disease. Having a disease-specific measure in pediatric GHD can be another useful clinical tool for monitoring patients to assess the impact of treatment, as well as to facilitate healthcare provider–patient communication. The measure is brief and covers three broad domains in areas that are typically not extensively covered in routine clinical practice. From a practical perspective, the measure could be completed by the parent/guardian after the patient has been checked in and waiting for the provider. A simple score could be calculated and incorporated into the patient's medical record, and then, after a period, the measure could be repeated. That would provide the clinician two key data points on GH therapy: (1) annualized height velocity as a primary endpoint, and (2) quality of life (QOL) data; both are important in assessing a patient's response to therapy. In addition, there are research implications. As new long-acting GH therapies are currently in clinical trials, a QOL measure would also serve as additional clinical data. Additionally, the GHD-CIM is intended to be used in research to assess the impact of new therapies and better understand the burden of disease.

## 5 Conclusion

The GHD-CIM ObsRO is a well-validated, adjuvant tool to assess disease-specific functioning that is currently not being adequately evaluated in research or clinical practice, providing a more complete patient-centric picture to the GH therapy experience.

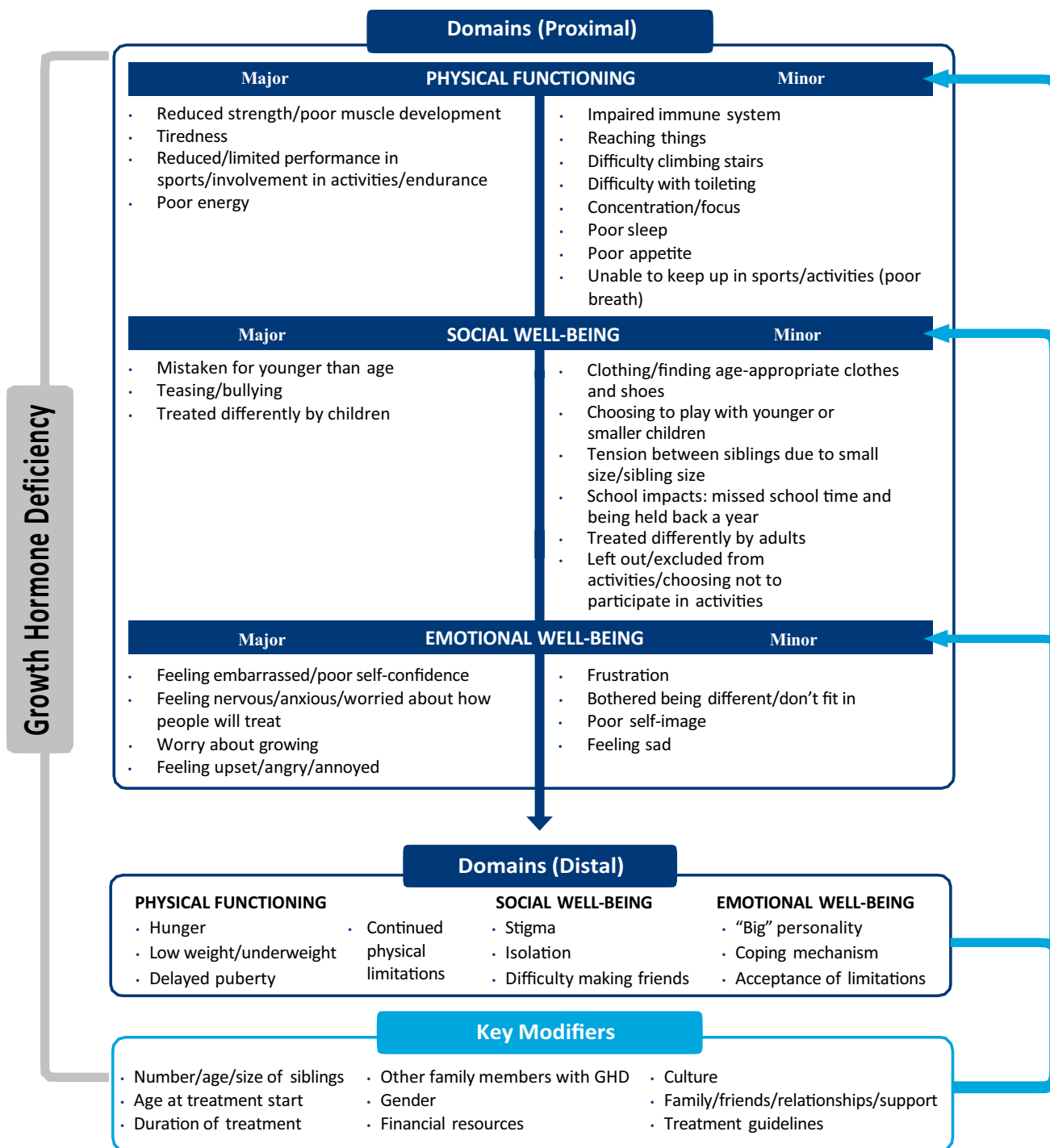


Fig. 3 GHD-CIM theoretical model. *GHD-CIM* Growth Hormone Deficiency-Child Impact Measure, *GHD* growth hormone deficiency

**Fig. 4** GHD-CIM conceptual framework. *GHD-CIM* Growth Hormone Deficiency-Child Impact Measure

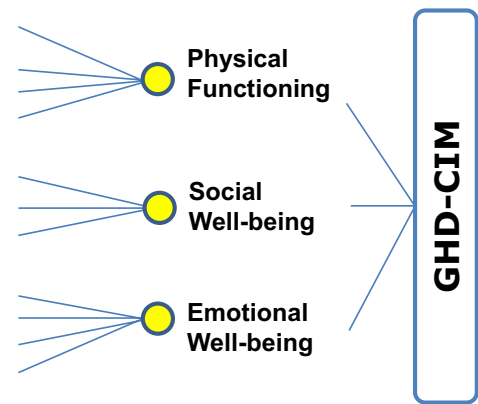
- How physically strong was child when doing sports or other physical activities
- How tired was child during the day
- How active was child
- How much energy did child have

Because of their size, how often:

- Did people think child was younger
- Were they teased
- Were they treated differently by children

Because of their size, how often child feel:

- Embarrassed
- Nervous or anxious
- Worried about growing
- Upset



**Acknowledgements** This research was supported by Novo Nordisk A/S. The authors thank Kai Wai Lee (Novo Nordisk) for his review and input into the manuscript. The research team would like to thank the clinical site Principal Investigators and research staff, and the study participants who generously shared their time and efforts for the purpose of this research study. *United States:* M. Jennifer Abuzzahab, MD, Children's Minnesota, MN; Kathleen Bethin, MD, PhD, University at Buffalo, NY; Samuel Casella, MD, Dartmouth-Hitchcock Medical Center, NH; Mariano Castro-Magana, MD, Winthrop-University Hospital, NY; Larry C. Deeb, MD, Tallahassee Memorial Healthcare, FL; Myrto Frangos, MD, Children's Endocrine of St Louis, MO; John Fuqua, MD, Riley Hospital for Children at Indiana University Health, IN; Luis Gonzalez-Mendoza, MD, Nicklaus Children's Hospital, FL; Evan Graber, DO, Nemours/Alfred I. duPont Hospital for Children, DE; Madeleine Harbison, MD, Mabel Yau, MD, and Ahmed Khat-tab, MD, Icahn School of Medicine at Mount Sinai, NY; Mahmoud Kabbani, MD, Pediatric Endocrine and Diabetes Clinic PC, AZ; Suzanne Kingery, MD, University of Louisville, KY; Peter Lee, MD, Penn State Health Milton S. Hershey Medical Center, PA; Aristides Maniatis, MD, Rocky Mountain Pediatric Endocrinology, CO; Rose Marino, MD, Massachusetts General Hospital, MA; Beth Motzkin-Kava, MD, Pediatric Endocrinology Consultants, FL; Linda Riddick, MD, Albany Medical College, NY; Douglas Rogers, MD, Cleveland Clinic, OH; Deborah Rotenstein, MD, AHN Pediatrics – Pediatric Alliance, PA; Ellen Sher, MD, TX; Robert Schultz, MD (deceased), Pediatric Endocrine Associates, GA; Kathryn Sumpter, MD (previous affiliation) and Grace Tannin, MD, UT Southwestern Medical Center, TX; Eva Tsalikian, MD, University of Iowa Hospitals and Clinics, IA; Moshe Weiss, MD, NY; W. Patrick Zeller, MD, Academic Endocrine Metabolism and Nutrition, Inc., IL; and William Zipf, MD, Central Ohio Pediatric Endocrine and Diabetes Services, OH. *United Kingdom:* Joanne Blair, MD, MCRP, MRCPC, Alder Hey Children's Hospital, Liverpool; Peter Clayton, MD, MRCP, FRCPC, Royal Manchester Children's Hospital, Manchester; Mehul Dattani, MD, FRCPC, FRCPC, Great Ormond Street Hospital for Children, London; and Jeremy Kirk, MD, FRCPC, FRCPC, DCH, Birmingham Women's and Children's Hospital, Birmingham.

**Code availability** Not applicable.

## Declarations

**Funding** This research was supported by Novo Nordisk.

**Conflict of interest** Meryl Brod, Suzanne Alopga, and Donald M. Bushnell are paid consultants to the pharmaceutical industry, including Novo Nordisk. Michael Højby Rasmussen, Knud Vad, and Jacques

Bedoin are employees of Novo Nordisk, A/S. Aristides Maniatis is a principal investigator for Novo Nordisk, Ascendis, Pfizer/OPKO, Genentech, and Sandoz.

**Ethics approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (Copernicus Group IRB, tracking number TBG1-15-428) and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

**Consent to participate** Informed consent was obtained from all respondents who participated in the study.

**Consent for publication** Not applicable.

**Availability of data and material** The data for the research presented in this publication may be available on a case-by-case basis upon reasonable request from the corresponding author.

**Author contributions** MB, MHR, KV, SA, and JB contributed to the study design and analysis and participated in manuscript preparation. DB conducted the analysis and participated in manuscript preparation. AM contributed to manuscript preparation and intellectual content. All authors read and approved the final version of the manuscript.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

## References

1. Donaldson MDC, Gregory JW, Van Vliet G, Wolfsdorf JI, Deladoey J. Short stature. In: practical endocrinology and diabetes in children. 2019. <https://doi.org/10.1002/9781119386230>. Accessed 22 Jan 2020.



2. Thomas M, Massa G, Craen M, de Zegher F, Bourguignon JP, Heinrichs C, et al. Prevalence and demographic features of childhood growth hormone deficiency in Belgium during the period 1986–2001. *Eur J Endocrinol.* 2004;151(1):67–72. <https://doi.org/10.1530/eje.0.1510067>.
3. Vimpani GV, Vimpani AF, Lidgard GP, Cameron EH, Farquhar JW. Prevalence of severe growth hormone deficiency. *Br Med J.* 1977;2(6084):427–30. <https://doi.org/10.1136/bmj.2.6084.427>.
4. Lindsay R, Feldkamp M, Harris D, Robertson J, Rallison M. Utah Growth Study: growth standards and the prevalence of growth hormone deficiency. *J Pediatr.* 1994;125(1):29–35. [https://doi.org/10.1016/s0022-3476\(94\)70117-2](https://doi.org/10.1016/s0022-3476(94)70117-2).
5. Brod M, Alolga SL, Beck JF, Wilkinson L, Hojbjerg L, Rasmussen MH. Understanding burden of illness for child growth hormone deficiency. *Qual Life Res.* 2017;26(7):1673–86. <https://doi.org/10.1007/s11136-017-1529-1>.
6. 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. *Horm Res Paediatr.* 2016;86(6):361–97. <https://doi.org/10.1159/000452150>.
7. Souza FM, Collett-Solberg PF. Adverse effects of growth hormone replacement therapy in children. *Arq Bras Endocrinol Metabol.* 2011;55(8):559–65. <https://doi.org/10.1590/s0004-27302011000800009>.
8. US FDA. Patient-reported outcome measures: use in medical product development to support labeling claims: guidance for industry. 2009. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims>. Accessed 20 Feb 2019.
9. European Medicines Agency. Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. 2005. [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2009/09/WC500003637.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003637.pdf). Accessed 20 Feb 2019.
10. Matza LS, Patrick DL, Riley AW, Alexander JJ, Rajmil L, Pleil AM, et al. Pediatric patient-reported outcome instruments for research to support medical product labeling: report of the ISPOR PRO good research practices for the assessment of children and adolescents task force. *Value Health.* 2013;16(4):461–79. <https://doi.org/10.1016/j.jval.2013.04.004>.
11. World Medical Association (WMA). WMA Declaration of Helsinki—Ethical principles for medical research involving human subjects. 2013. <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>. Accessed 10 May 2019.
12. International Society for Pharmacoepidemiology. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol Drug Saf.* 2008;17(2):200–8. <https://doi.org/10.1002/pds.1471>.
13. US Office of the Federal Register. Electronic Code of Federal Regulations. Titles 21§50 Protection of human subjects and §56 Institutional review boards. Washington, DC: Office of the Federal Register; 2018. [https://www.ecfr.gov/cgi-bin/text-idx?SID=25670b7f62f18a7a8771e34b0ece4cbe&mc=true&tpl=/ecfrbrowse/Title21/21cfrv1\\_02.tpl#0](https://www.ecfr.gov/cgi-bin/text-idx?SID=25670b7f62f18a7a8771e34b0ece4cbe&mc=true&tpl=/ecfrbrowse/Title21/21cfrv1_02.tpl#0). Accessed 10 May 2019.
14. Prader A, Largo RH, Molinari L, Issler C. Physical growth of Swiss children from birth to 20 years of age. First Zurich longitudinal study of growth and development. *Helv Paediatr Acta Suppl.* 1989;52:1–125.
15. Bullinger M, Quitmann J, Power M, Herdman M, Mimoun E, DeBusk K, et al. Assessing the quality of life of health-referred children and adolescents with short stature: development and psychometric testing of the QoLISSY instrument. *Health Qual Life Outcomes.* 2013;11:76. <https://doi.org/10.1186/1477-7525-11-76>.
16. Bullinger M, Schmidt S, Petersen C, Disabkids Group. Assessing quality of life of children with chronic health conditions and disabilities: a European approach. *Int J Rehabil Res.* 2002;25(3):197–206. <https://doi.org/10.1097/00004356-200209000-00005>.
17. Whiteside SP. Adapting the Sheehan disability scale to assess child and parent impairment related to childhood anxiety disorders. *J Clin Child Adolesc Psychol.* 2009;38(5):721–30. <https://doi.org/10.1080/15374410903103551>.
18. Snoek FJ, Mollema ED, Heine RJ, Bouter LM, van der Ploeg HM. Development and validation of the Diabetes Fear of Injecting and Self-Testing Questionnaire (D-FISQ): first findings. *Diabet Med.* 1997;14(10):871–6. [https://doi.org/10.1002/\(SICI\)1096-9136\(199710\)14:10%3c871:AID-DIA457%3e3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1096-9136(199710)14:10%3c871:AID-DIA457%3e3.0.CO;2-Y).
19. SPSS Inc. PASW Statistics for Windows. 18.0 ed. Chicago, IL: SPSS Inc.; 2009.
20. Carmines EC, Zeller RA. Reliability and validity assessment. Beverly Hills: Sage Publications; 1979.
21. Gliem JA, Gliem RR. Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. In: 2003 Midwest research to practice conference in adult, continuing, and community education; 2003. pp. 82–8.
22. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16(2):297–334.
23. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: L. Erlbaum Associates; 1988.
24. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc.* 2002;77(4):371–83. <https://doi.org/10.4065/77.4.371>.
25. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care.* 1999;37(5):469–78. <https://doi.org/10.1097/00005650-199905000-00006>.
26. Byrne BM. Structural equation modeling with AMOS: basic concepts, applications, and programming. New York: Routledge; 2001.
27. Arbuckle JL. AMOS 20.0 user's guide. Crawfordville: Amos Development Corporation; 2011.
28. Kline RB. Principles and practice of structural equation modeling. 3rd ed. New York: Guilford Publications; 2010.
29. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model.* 1999;6:1–55.
30. Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. New York: McGraw-Hill; 1994.