ORIGINAL RESEARCH ARTICLE

Psychometric Validation of the Growth Hormone Deficiency–Child Treatment Burden Measure (GHD-CTB) and the Growth Hormone Deficiency–Parent Treatment Burden Measure (GHD-PTB)

Meryl Brod[1] · Michael Højby Rasmussen[2] · Suzanne Alolga[1] · Jane F. Beck[1] · Donald M. Bushnell[3] · Kai Wai Lee[2,5] · Aristides Maniatis[4]

Abstract

Purpose The aim was to evaluate the measurement properties of the Growth Hormone Deficiency–Child Treatment Burden Measure–Child (GHD-CTB–Child), a patient-reported outcome (PRO) for children aged 9 to < 13 years; the Growth Hormone Deficiency–Child Treatment Burden Measure–Observer (GHD-CTB–Observer), an observer-reported outcome (ObsRO) version completed by parents/guardians of children with growth hormone deficiency (GHD) aged 4 to < 9 years; and the Growth Hormone Deficiency–Parent Treatment Burden Measure (GHD-PTB), a PRO that assesses the treatment burden of parents/guardians living with children with GHD aged 4 to < 13 years.

Methods A non-interventional, multi-center, clinic-based study across 30 private practice and large institutional sites in the United States and the United Kingdom was conducted. The sample consisted of 145 pre-pubertal children aged 9 to < 13 years at enrollment with a physician confirmed GHD diagnosis as well as 98 parents/guardians of pre-pubertal younger children aged 4 to < 9 years at enrollment with a physician confirmed GHD diagnosis. The child sample consisted of 59 treatment-naïve children (no prior exposure to growth hormone [GH] therapy; were starting GH treatment at study start per standard of care) and 184 children already maintained on treatment for at least 6 months. At baseline, all study participants completed a paper validation battery including all measures needed to conduct the validation analyses. Follow-up assessments with children in the maintenance group and their caregiver/parent were conducted approximately 2 weeks post-baseline to evaluate test–retest reproducibility. To evaluate sensitivity to change and meaningful change thresholds, treatment-naïve participants in both child and parent/guardian populations were assessed within 1 week of report of minimal improvement between week 3 and week 11 and at week 12. Psychometric analyses were implemented following an a priori statistical analysis plan.

Results Factor analyses confirmed the a priori conceptual domains and Overall score for each measure (GHD-CTB–Child and GHD-CTB–Observer domains: Physical, Emotional Well-being, and Interference; GHD-PTB domains: Emotional Well-being and Interference). Internal consistency was acceptable for all measures (Cronbach's alpha > 0.70). Test–retest reliability was acceptable for the Physical, Emotional, and Overall domains of the GHD-CTB versions, and the Emotional and Overall domains of the GHD-PTB (intraclass correlation coefficient above 0.70). All but one of the convergent validity hypotheses for the GHD-CTB versions and all hypotheses for the GHD-PTB were proven ($r > 0.40$). Known-groups validity hypotheses were significant for length of time to administer the injections in the GHD-CTB versions ($p < 0.001$ for Physical, Emotional, and Overall, and $p < 0.01$ for Interference) and whether parents/guardians versus child gave the injections more often for the Emotional domain of the GHD-PTB ($p < 0.05$). Associated effect sizes ranged from −0.27 to −0.57 for GHD-CTB versions and from −0.74 to −0.69 for the GHD-PTB, indicating that the measures are sensitive to change. Anchor-based patient and parent/guardian ratings of severity suggest preliminary meaningful change thresholds (GHD-CTB: 6 points for Physical score, 9 for Emotional, and 6 for Interference; GHD-PTB: 10 points for Emotional and 6 for Interference scores).

Conclusions The psychometric properties of the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB support the validity of their use as PRO and ObsRO measures to capture the experiences associated with treatment burden for children with GHD and their parents/guardians in both clinical and research settings.

The Clinicaltrials.gov registration number NCT02580032 was first posted October 20, 2015.

Extended author information available on the last page of the article

△ Adis

Key Points for Decision Makers

Evaluation of the Growth Hormone Deficiency–Child Treatment Burden Measure (GHD-CTB)–Child, GHD-CTB–Observer, and Growth Hormone Deficiency–Parent Treatment Burden Measure (GHD-PTB) showed them to be reliable, valid measures of treatment burden for children and adolescents with growth hormone deficiency and their parents/guardians.

The GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB can serve as useful tools in both clinical and research settings of the child's overall response to therapy.

1 Background

Growth hormone deficiency (GHD) is a condition that occurs when there is insufficient production of growth hormone (GH) from the pituitary gland. GHD has multiple etiologies, including idiopathic, genetic, injury related, or arising from another medical condition, e.g., brain tumor [1]. In children, GHD leads to decreased growth rates, below average height, low energy levels, impaired metabolism, and weakened muscle strength and stamina, as well as negative social and emotional impacts on quality of life (QOL), partially attributed to being smaller and looking younger than their peers [2–8]. Prevalence estimates of GHD in children range between 1.8 and 2.9 per 10,000 based on data from Europe and the United States (US) [9–11]. Research results from three large non-interventional, multi-center registry studies conducted in the US, Europe, and/or Japan found that prepubertal children were started on treatment for all causes of GHD at a median age of between 5.3 and 9.7 years (depending on GHD diagnosis and sex) in the KIGS registry (n = 83,803) and at an average age of 9.7 years in the ANSWER and Nordinet IOS studies (combined studies n = 37,702), and the children were predominately male (57–70.1%) [12–14].

GHD may be treated with recombinant GH replacement therapy [15, 16], which has been available for more than 35 years and shown to be safe overall [17–21]. The treatment regimen normally consists of an ongoing daily injection throughout childhood and adolescence until obtaining adult height [22]. With early diagnosis and treatment initiation, children with GHD often achieve a height within the normal range for their family; research further indicates beneficial impacts on their social and emotional well-being

[5, 7, 23–25]. In the US, a retrospective review of two claims databases for children diagnosed with GHD between 2007 and 2018 who were insured by either Medicaid (n = 6820) or commercial health insurance (n = 14,070) found 63.2% (Medicaid) and 68.4% (commercial) GHD patients were treated at some point with somatropin [26].

Treatment burden has been identified as a key driver of adherence to treatments [27, 28]. Poor adherence and long-term persistence rates have been identified as key areas of concern in children and adolescents receiving GH replacement therapy [29–40] as there is evidence that lower adherence is associated with lower growth rates [29–31, 41]. Additionally, there may be other factors that influence or modify the treatment burden experience. The literature suggests that children who start GH treatment at an earlier age will have fewer impacts (e.g., emotional) than children who start at an older age [42, 43]; children who have been on treatment longer will have less physical pain [42]; the longer it takes to prepare and administer the injection, the greater emotional burden on the child [44]; parents/guardians who have greater responsibility for disease management including giving injections more often than their child self-injects will have greater parent/guardian emotional burden [45, 46]; and parents/guardians of older children who have lesser responsibility for disease management (e.g., child self-injects) will have less interference than parents/guardians of younger children for whom they have greater responsibility for disease management) [e.g., including administering the injections] [47]. Thus, understanding and addressing treatment burden for children with GHD and their parents/guardians is critical to improving adherence and persistence to optimize treatment outcomes.

Unfortunately, although there are generic or non-GHD-specific measures available that can be used to assess treatment burden in GHD, no well-validated, GHD-disease-specific measures exist that comprehensively examine the issues associated with the burden of GH treatment for children with GHD and their parents. Disease-specific measures allow for greater sensitivity to concepts of interest and are more responsive to changes in status over time [48, 49]. To assess treatment burden associated with GHD, two disease-specific measures were developed to assess treatment burden for children with GHD as well as treatment burden for the parents of these children. The measures were developed according to the US Food and Drug Administration (FDA) guidance on patient-reported outcome (PRO) measure development [50]. The concept elicitation phase included a review of the literature, interviews with clinical experts, and four focus groups (in Germany) and 52 telephone interviews in the United Kingdom (UK) and US, which were conducted with children/adolescents with GHD aged 8 to < 13 years and parents of children with GHD aged ≥ 4 to

< 13 years. Analysis identified 3 major areas of GHD treatment burden for children: physical, emotional well-being, and interference. Based on this information, items for the preliminary measures were developed, cognitively debriefed in 13 children and 13 parents, and validation-ready versions of the measures were developed.

Specifically, these measures are as follows:

1. The Growth Hormone Deficiency–Child Treatment Burden Measure (GHD-CTB).

(a) *GHD-CTB–Child*: A PRO assessing the treatment burden of GHD on children aged 9 to < 13 years. The measure has 14 items, with domains of Physical, Emotional Well-being, and Interference, with response options for how often, ranging from 'Never' to 'All of the time,' and response options for how much, ranging from either 'Not at all' or 'None' to 'Extreme.'

(b) *GHD-CTB–Observer*: Assesses treatment burden for the child using an observer-reported outcome (ObsRO) version of the GHD-CTB–Child completed by parents/guardians living with children with GHD aged 4 to < 9 years. The 14 items in the ObsRO version reflect the same content and domains as the PRO version with the following instructions: "When answering the questions, please check the response box that most closely represents what you have SEEN or BEEN TOLD by your child or by others about your child. If you have not seen or been told anything which informs you how to answer a question, please check the 'Don't Know' response box. Please do not answer any questions based on what you think, base your response only on what you have seen or been told."

Since the measures were intended to be used in clinical trials of GH treatments, the age range of the sample used to develop the measures was selected to closely match the eligibility requirements of the trials for which they would be used in as per FDA guidelines [50]. Additionally, the lower age (4 years) was suggested by literature on age of GH treatment initiation [12, 13, 51]. The appropriate age ranges for PRO versus ObsRO versions were determined based on the findings from the concept elicitation phase of the measure development and literature suggesting appropriate ages for children to answer PRO measures [44]. Children under the age of 8 years are generally not able to answer questions of this type on their own [52]. Therefore, in the concept elicitation phase, interviews were conducted with children starting at age 8. However, the lower age for the PRO version was set at age 9 years as children younger than that had issues with comprehension and interpretation of the recall period, were less able to complete, and reported more nervousness

than the older children. The full methods and results for the conceptual development study phase have been previously reported [44].

2. The Growth Hormone Deficiency–Parent Treatment Burden Measure (GHD-PTB).

GHD-PTB: A parent/guardian PRO assessing the treatment burden of parents/guardians of children aged 4 to < 13 years. The measure has eight items, with domains for Emotional Well-being and Interference, with response options for how much, ranging from 'Not at all' to 'Extremely,' and for how often, ranging from 'Never' to 'All of the time.' This measure assesses burden from the perspective of parents about themselves.

This article describes the psychometric validation data for all three measure versions.

2 Methods

Sound scientific principles of PRO and ObsRO measure development were followed according to FDA and European Medicines Agency regulatory guidance, and guidance provided by the International Society for Pharmacoeconomics and Outcomes Research for pediatric PRO development [50, 53, 54]. In the US, ethics approval was obtained from Copernicus Group Independent Review Board (Tracking #: 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 [55] were followed, and signed consent/assent was obtained before initiating study-related activities.

2.1 Data Management

Quality assurance (QA) review was conducted, and all forms were edited for clarity and to correct, or flag, data inconsistencies and/or missing data. If there were any inconsistencies found on the surveys, editing rules were used to determine what data to enter. For example, if two or more responses on a scale or numeric range were chosen for the same question, then the mid-point, or average of the numbers, was used.

All data were entered by the same person. And a second QA check was conducted at data entry. Any unresolved data issues were marked as missing. Verification was also done post data entry on all surveys by a visual comparison of the paper survey against the database. Data entry errors and corrections were updated in the database and noted in a change log.

2.2 Validation Phase

The validation phase of the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB was carried out concurrently as a joint study with the validation of the Growth Hormone Deficiency–Child Impact Measure (GHD-CIM). The full study design, eligibility criteria, assessments, and statistical analysis methods of relevance to both measures have been described in detail elsewhere [56]. A summary is provided in the subsections below.

2.2.1 Study Design

A non-interventional, multi-center, clinic-based study was carried out across 30 private practice and large institutional sites in the US and UK.

2.2.2 Study Sample

The study sample included 145 pre-pubertal children aged 9 to < 13 years at enrollment with a physician-confirmed GHD diagnosis who answered questions independently (the GHD-CTB–Child and other validation battery assessments), their parents/guardians (*n* = 145) who completed the parent treatment burden PRO (GHD-PTB), and 98 parents/guardians of pre-pubertal younger children aged 4 to < 9 years at enrollment with a physician-confirmed GHD diagnosis who both completed the GHD-PTB and also answered questions as an observer (GHD-CTB–Observer and other validation assessments) for children not able to answer for themselves. Thus, the total sample for the GHD-CTB–Child was *N* = 145, for the GHD-CTB–Observer, it was *N* = 98, and for the GHD-PTB, it was *N* = 243. Fifty-nine children in the sample were, at enrollment, treatment naïve (no prior exposure to GH therapy; were starting GH treatment at study start per standard of care) and 184 of the sample were, at enrollment, on maintenance (currently on GHD treatment, for at least 6 months). Maintenance child participants were treated with any commercially available product per the standard of care, with no study interventions.

Sample size calculations were based on numbers needed to achieve adequate validation results based on factor analysis, and generally, five people per item are needed [57, 58]. While the pre-validation GHD-CTB contained 17 items, the sample size was based on the 33-item GHD-CIM also evaluated in this study. Based on this, a sample size of at least 165 was needed. To ensure an adequate sample size, which also takes into consideration the reality that treatment-naïve participants are difficult to recruit given that GHD is a rare condition and that the study inclusion criteria are strict to ensure a true GHD population, we defined a minimum of 200 and maximum of 320 participants with complete information to be sufficient to establish validity, and reliability, and to

conduct a preliminary examination of sensitivity to change and assessment of meaningful change threshold (MCT).

2.2.3 Validation Study Visits and Assessments

At baseline, study participants completed a paper validation battery (see Table S1, Online Resource 1, in the Electronic Supplementary Material, for a description of measures used in the validation study), which included socio-demographic items, relevant medical history, the GHD-CTB–Child or GHD-CTB–Observer, the GHD-PTB, and the one-item Patient Global Impression of Severity (PGIS) [the PGIS is rated on a 6-point scale, with the severity of illness response scale being 1—'No noticeable symptoms,' 2—'Very mild,' 3—'Mild,' 4—'Moderate,' 5—'Severe,' and 6—'Very severe']. The battery also included the QoLISSY [59], DIS-ABKIDS (DCGM-37) [60], Child Sheehan Disability Scale (CSDS) [61], and Diabetes Fear of Injecting and Self-Testing Questionnaire (D-FISQ) [62] /Fear of Self-Injecting (FSI) subscale. Overall treatment burden items tied to each of the domains of the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB were also developed and included in the battery to aid in the assessment of construct validity. Clinicians completed the Clinician Global Impression of Severity (CGIS) [CGIS is rated on a 6-point scale, with the severity of illness response scale being 1—'Very severe,' 2—'Severe,' 3—'Moderate,' 4—'Mild,' 5—'Very mild,' and 6—'No noticeable symptoms'] and clinical measurements.

Follow-up assessments were conducted approximately 2 weeks post-baseline to evaluate test–retest reproducibility in the Maintenance group. Items to assess change within the retest period were included: 'have you (has your child) experienced any major life events since the last study visit' and 'have the past 2 weeks been an unusually stressful period for you (your child)' (these items are called the Changes Since Last Assessment questionnaire). The parent answered these questions about themselves and about the child. To evaluate sensitivity to change and MCTs of the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB, treatment-naïve participants in both child and parent/guardian populations were assessed within 1 week of report of minimal improvement between week 3 and week 11 and at week 12 using a two-question Patient Global Rating of Change scale, with the first asking the participant whether their GHD condition has stayed the same, gotten better or worsened, and then asking how much better (on a 6-point scale, ranging from 'Almost the same, hardly better at all' to 'A very great deal better') or how much worse (also on a 6-point scale, ranging from 'Almost the same, hardly worse at all' to 'A very great deal worse').

The GHD-CTB and GHD-PTB are scored by summing across the items to compute a raw score and converting it to a 0- to 100-point standardize score, with a higher score

indicating a higher treatment burden. All three domains (Physical, Emotional, and Interference) of the GHD-CTB–Child and GHD-CTB–Observer and both domains (Emotional and Interference) of the GHD-PTB must be scored in order to compute a total score for each measure. Up to one missing item per domain is allowed, otherwise, that domain score and, thus, the total score are not generated. The ObsRO version includes a response option of 'Don't know.' For scoring purposes, this response was coded as missing. The Overall score is calculated as the mean of the domain scores.

2.3 Statistical Analysis Methods

Analyses were carried out in accordance with an a priori statistical analysis plan and conducted using SPSS [63]. A significance test level of $p < 0.05$ (two-sided) was used unless otherwise noted. Statistical tests involving multiple comparisons (e.g., analysis of variance [ANOVA] models with multiple groups) included Scheffe post hoc tests.

2.3.1 Sociodemographic and Clinical Characteristics

Sociodemographic and clinical characteristics were calculated using descriptive statistics.

2.3.2 Evaluation of Measure Items

Evaluation of the items in the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB was made using information from the following analyses:

- *Floor and ceiling effects*: Descriptive statistics (*N*, frequency distribution, mean, median, range, standard deviation [SD], percentage at floor, percentage at ceiling) were calculated for the individual item responses for the GHD-CTB and GHD-PTB items. The ceiling effect (responses at the least severe end of the scale, i.e., where participants cannot get any better) threshold for closer examination was set at 50%.
- *Item-to-item correlations*: Item-to-item correlation is a Pearson's correlation matrix of each item in the questionnaire. To determine the item-to-item correlation, a reliability analysis was conducted for all item pairs. A focus was made on correlation coefficients of greater than 0.50 indicating potential redundancy between the items [64].
- *Item-to-total correlations*: Item-to-total correlation is a Pearson's correlation between each item score and the total score. The item score is the individual score for each item in the GHD measures, and the total score is the summation of all the items per GHD measure. To calculate the item-to-total correlation, 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 less than 0.40 [65] was examined, since this indicates that it may not be sufficiently associated with the remaining items in the hypothesized scale.

- *Exploratory and confirmatory factor analyses (CFA)*: Exploratory factor analysis procedures were performed on the correlation matrices derived from the items comprising the GHD measures. Factor analysis provides a means of analyzing the relationships among inter-correlations of the items. Rotational methods 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. The results of the factor analysis were used to guide the development of the scoring algorithm for the GHD-CTB and GHD-PTB measures. A post hoc CFA was also performed on the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB using IBM® SPSS® Amos™ Version 20 (2019) [66, 67]. 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.
- *Internal consistency reliability*: Cronbach's alpha was used to assess internal consistency reliability [68]. 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 will be necessary to claim the instrument is internally consistent, and it is preferred to have alpha values between 0.80 and 0.90.
- *Test–retest reliability*: Intraclass correlation coefficient (ICC) [two-way mixed model with absolute agreement] was used to assess test–retest reliability in a subsample from the Maintenance group who were administered the retest approximately 2 weeks after baseline and indicated experiencing no change on the Changes Since Last Assessment items (major life events or treatment).
- *Construct convergent validity:* Convergent construct validity, comparing the measure to other logically related measures, was conducted [69]. Pearson's correlation was computed to measure the association between the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB scores and the other measures included in the study. Convergent validity was supported when the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB scores were significantly correlated with items or instruments measuring similar concepts. An $r \geq 0.40$ was considered acceptable, noting

that a stronger correlation (> 0.50) offers greater support for the relationship [70]. The hypotheses tested are listed in Table S2 (see Online Resource 1 in the Electronic Supplementary Material).

- *Known-groups validity*: Known-groups validity is defined as the ability of a measure to differentiate between independent groups known to differ. In order to test whether the sample was greater than or less than the desired range of values (testing both ends/tails of the normal distribution), a two-tailed test at the $p < 0.05$ level was used for testing the hypotheses. (When more than one hypothesis per subdomain was proposed, an a priori decision was made that proving one hypothesis per domain would be sufficient to claim validity.) Hypotheses were derived based on clinical experience and data available in the literature. The a priori hypotheses tested are listed in Table S3 (see Online Resource 1).

- *Sensitivity to change*: An exploratory analysis of potential sensitivity to change was conducted using distributional methods to evaluate the effect size (mean change divided by the baseline SD). Higher values for the effect size indicated a greater sensitivity to change, with values of 0.2–0.5 regarded as 'small,' 0.5–0.8 as 'moderate,' and those above 0.8 as 'large' [71].

- *Interpretation of meaningful change*: Anchor-based methods were used to examine meaningful within-patient change, 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. Our purpose was to identify what was the smallest yet meaningful change to patients. Meaningful change was defined as the difference between these two momentary assessments of GHD severity (baseline and 12-week follow-up), with differences anchored to changes in one response option (e.g., 'Severe' to 'Moderate') or two response options (e.g., 'Severe' to 'Mild') [72]. Each of the meaningful change differences derived using the various methods were examined with the goal of converging to a final single estimate by triangulating these differences (i.e., averaging between the various values). Thus, the final single estimate took into consideration both what would be considered as beneficial from the patient's viewpoint and clinically meaningful.

3 Results

A total of 252 individuals were enrolled. Eight individuals screen-failed, and one was withdrawn for protocol violation. The remaining 243 individuals (145 self-reporting children aged 9 to < 13 years with 145 of their parents/guardians, and 98 parents/guardians of non-self-reporting children aged 4 to < 9 years) were included in the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB item-level analyses and convergent validity analyses.

3.1 Statistical Analysis

3.1.1 Sociodemographic and Clinical Characteristics

Children with GHD were predominantly from the US (91.8%), male (71.9%), and white (84.8%), with a mean age of 9.2 years (range 4–13 years of age). The child's mean age at diagnosis ranged between 4.6 years (parent–maintenance group) and 10.7 years (child–treatment-naïve group) [total mean, 6.9 years]. Similarly, the child's mean age when they first started GH therapy ranged from 4.8 years (parent–maintenance group) to 10.8 years (child–treatment-naïve group) [total mean, 7.1 years]. The majority (78.6%) used a pen as their medication injection device, with about 4.9% using needle and syringe. Less than a quarter of the children (18.1%) had been prescribed other medications. Other health conditions included ear, nose, and throat (10.5%), mental health (10.1%), respiratory (9.7%), and endocrine disorders (8.9%); over half (54.0%) indicated having no other health conditions.

Parent/guardian mean age was 41.6 years (range 25–66), with parents/guardians of children who self-reported being, on average, 5 years older than parents/guardians who completed the observer assessments. Most were from the US (91.8%), mothers (80.7%), and married (88.1%); about half (51.0%) worked full-time for pay, with 23.0% not working due to other reasons (not retired or disabled). The demographic and clinical characteristics of child and parent/guardian participants are shown in Tables S4 and S5 (see Online Resource 1 in the Electronic Supplementary Material).

3.1.2 Item Reduction

For both the GHD-CTB–Child and GHD-CTB–Observer and for the GHD-PTB, there were several pairs of items that had item-to-item correlations of greater than 0.50, indicating potential redundancy. These were examined closely for item reduction. Three items were removed from the GHD-CTB–Child and GHD-CTB–Observer and four from the GHD-PTB after being deemed to be conceptually redundant. The final GHD-CTB–Child and GHD-CTB–Observer have 14 items, and the final GHD-PTB has eight items.

Two of the conceptually redundant items removed from the GHD-CTB–Child and GHD-CTB–Observer also had high ceiling effects. However, the remaining items with high ceiling effects were retained because they had previously been confirmed as both relevant and important by respondents in the cognitive debriefing interviews conducted during

the development study phase. See Table 1 for the item reduction tracking table.

Following item reduction, the remainder of the psychometric analyses were conducted.

3.1.3 Descriptive Characteristics of GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB Items

For retained items in the final measures, mean scores for each of the 14 items of the GHD-CTB–Child and GHD-CTB–Observer ranged from 0.17 to 1.54 using a response scale between 0 'Not at all/None/Never' and 4 'Extremely/All of the time.' Eleven out of 14 items used the full range (0, 1, 2, 3, and 4) of responses. Some items exhibited a ceiling effect where at least 50% of respondents used either the 'Not at all' or 'Extremely' response. Consequently, the means and medians were lower than expected. Missing data were minimal (0.7–5.1%). The GHD-CTB–Observer included a response option of 'Don't know.' For retained items, these responses ranged from 2.0 to 8.2%, with the higher report for 'How much soreness did you have in places on your body where you got your shots?'

Item-to-total correlations, for most GHD-CTB–Child and GHD-CTB–Observer items, showed a strong association between each item against the rest of the items as a total score (excluding that item). Associations lower than 0.40 were 'How often were you worried about remembering to take your shots?' (0.319), 'How often did you miss doing things because of your shots?' (0.303), and 'How often did your shots interrupt or get in the way of the things you wanted to do?' (0.360). All other correlations were above 0.40.

The examination of the item characteristic data supported that there were not substantial differences between what parents and children reported, which would have required examining the groups independently, and, thus, the validation was conducted on the pooled data set.

Mean scores for each of the retained eight items of the GHD-PTB ranged from 0.35 to 1.28 using a response scale between 0 'Not at all/Never' and 4 'Extremely/All of the time'. All items used the full range of responses except the item 'How often did your child's treatment interfere with your social life?', which did not have a response of 4 'All of the time.' As seen with the GHD-CTB–Child and GHD-CTB–Observer, the overall trend was toward the 'better' end of the scale (responses of 'Not at all/Never' or 'A little/Rarely'). The majority of items exhibited a ceiling effect (five of eight items were over 50%). Missing data were minimal (2.9–3.3%).

Item-to-total correlations, for most GHD-PTB items, showed a strong association between each item against the rest of the items as a total score (excluding that item). Associations lower than 0.40 were all within the

Interference domain with the item stem 'How often did your child's treatment interfere with your social life?' (0.349) and 'travel plans?' (0.274). All other correlations were above 0.40.

The descriptive statistics for retained individual item responses on the GHD-CTB–Child and GHD-CTB–Observer are provided in Tables S6 and S7 (see Online Resource 1 in the Electronic Supplementary Material). The descriptive statistics for individual item responses on the GHD-PTB are shown in Table S8 (see Online Resource 1).

3.1.4 Factor Analyses

Three factors are presented for the GHD-CTB–Child and GHD-CTB–Observer and two for the GHD-PTB (Table 2). Exploratory factor analyses confirmed the a priori conceptual framework for each measure. All items comprising the Emotional and Interference domains of each measure, and the Physical domain of the GHD-CTB–Child and GHD-CTB–Observer, factored into those domains. One item, 'How much did your shots hurt?', had coefficients that would allow that item to be in either Emotional (0.610) or Physical (0.512). It was retained in the Physical domain as it was hypothesized as a physical symptom.

CFA results confirmed adequate fit indices: comparative fit index (0.967 for the GHD-CTB–Child and GHD-CTB–Observer and 0.946 for the GHD-PTB), goodness-of-fit index (0.987 for the GHD-CTB–Child and GHD-CTB–Observer and 0.964 for the GHD-PTB), root mean square residual (0.0518 for the GHD-CTB–Child and GHD-CTB–Observer and 0.064 for the GHD-PTB), and root mean square error of approximation (0.08 for the GHD-CTB–Child and GHD-CTB–Observer and 0.09 for the GHD-PTB) [73, 74].

Additionally, a higher order factor analysis was conducted on the three subscale scores of the GHD-CTB–Child and GHD-CTB–Observer and the two subscale scores of the GHD-PTB to determine the ability to create an overall score of treatment burden. The subscales factored into a single component, with 62.9% of total variance explained for the GHD-CTB–Child and GHD-CTB–Observer and 64.6% for the GHD-PTB.

3.1.5 Reliability

Internal consistency reliability was examined and resulted in a Cronbach's alpha of 0.875 for the GHD-CTB–Child and GHD-CTB–Observer Overall score and 0.745 for the GHD-PTB Overall score; all coefficients by domain also exceeded the threshold of 0.70, indicating internally consistent scales for both measures, as shown in Tables S9 and S10 (see Online Resource 1 in the Electronic Supplementary Material).

Table 1 Item reduction table

	Keep/drop	Rationale for dropping
GHD-CTB items		
CTB01 How much shots hurt	Keep	
CTB02 How much shots sting	Keep	
CTB03 How much shots burn	Drop	Correlated with 02 'Sting' (0.70), high ceiling effect (70%)
CTB04 How much bruising	Keep	
CTB05 How much soreness	Keep	
CTB06 How often worried about getting shots	Keep	
CTB07 How often worried about shots hurting	Keep	
CTB08 How often worried about remembering to take	Keep	
CTB09 How often embarrassed take shots to grow	Drop	Correlated with 13 'Diff needing shots' (0.55), high ceiling effect (73%)
CTB10 How often scared about getting shots	Keep	
CTB11 How often annoyed by shots	Keep	
CTB12 How often unhappy about how often take shots	Keep	
CTB13 How often feel bothered different	Keep	
CTB14 How often try to avoid/delay getting shots	Drop	Correlated with 06 'Worried' (0.60) and 10 'Scared' (0.61)
CTB15 How often miss doing things because of shots	Keep	
CTB16 How often need to change plans because of shots	Keep	
CTB17 How often shots interrupt wanted to do	Keep	
GHD-PTB items		
PTB18 Worry causing your child pain when giving the injection	Keep	
PTB19 Worry remembering to give the injection	Keep	
PTB20 Worry doing the injection correctly	Keep	
PTB21 Worry preparing the injection correctly	Drop	Correlated with 20 'Doing injection correctly' (0.77), high ceiling effect (73%)
PTB22 Often anxious about giving the injection	Drop	Correlated with 20 'Doing injection correctly' (0.79) and 18 'Causing pain when injecting' (0.69), high ceiling effect (61%)
PTB23 Often guilty that your child needs injections	Drop	Correlated with 25 'Sad child needs injections' (0.77)
PTB24 Often frustrated with your child's behavior related to injections	Keep	
PTB25 Often sad about your child needing injections	Keep	
PTB26 Often interfere with your social life	Keep	
PTB27 Often interfere with your travel plans	Keep	
PTB28 Often interfere with your daily routine	Keep	
PTB29 Often interfere with your family schedules	Drop	Correlated with 28 'Interfere with daily routine' (0.81), high ceiling effect (70%), low item-to-total correlation (0.39)

CTB Child Treatment Burden, *GHD-CTB* Growth Hormone Deficiency–Child Treatment Burden Measure, *GHD-PTB* Growth Hormone Deficiency–Parent Treatment Burden Measure, *PTB* Parent Treatment Burden

Test–retest reproducibility was assessed using the ICC in a subsample from the maintenance group who indicated experiencing no change on the Changes Since Last Assessment items. For the GHD-CTB–Child and GHD-CTB–Observer, the ICCs were adequate for the Physical (0.76), Emotional (0.78), and Overall domains (0.81), but below the threshold of 0.70 for Interference (0.64). For the GHD-PTB, the ICCs were also adequate for the Emotional (0.806) and Overall domains (0.783), but below the threshold for Interference (0.602).

In order to ensure that it was appropriate to combine both child and parent observer data for the GHD-CTB, reliability for both groups was also examined. Results confirmed internal consistency (Cronbach's alpha and test–retest reproducibility) were similar, supporting combining data for both groups (see Online Resource 1, Table S11).

3.1.6 Convergent Validity

For the GHD-CTB–Child and GHD-CTB–Observer, convergent validity was assessed by examining the magnitude of correlations between the GHD-CTB–Child and GHD-CTB–Observer scores and the DISABKIDS, QoLISSY, CSDS, and individual treatment burden items. All but one (six out of seven) hypothesized association had statistically significant correlations over 0.40, with four of these being greater than 0.50. The GHD-CTB Overall score with QoLISSY Treatment domain score was significant ($p < 0.01$), with a correlation of 0.34. Significant correlations were found for the GHD-CTB–Child and GHD-CTB–Observer Overall with the overall treatment burden item (0.58, $p < 0.001$), GHD-CTB–Child and GHD-CTB–Observer Physical with the overall treatment burden item (0.46, $p < 0.001$), GHD-CTB–Child and GHD-CTB–Observer Emotional with the DISABKIDS Treatment domain (0.72, $p < 0.001$), GHD-CTB−Child and GHD-CTB–Observer Emotional with overall treatment burden item (0.42, $p < 0.001$),

GHD-CTB−Child and GHD-CTB–Observer Child Interference with the CSDS Total score (0.55, $p < 0.001$), GHD-CTB−Child and GHD-CTB–Observer Child Interference with the CSDS proxy score (CSDS-P) [0.44, $p < 0.001$], and GHD-CTB−Child and GHD-CTB–Observer Child Interference with overall interference of treatment item (0.60, $p < 0.001$). See Table S2 (Online Resource 1 in the Electronic Supplementary Material) for actual correlation coefficients and significance levels.

For the GHD-PTB, convergent validity was assessed by examining the magnitude of correlations between the GHD-PTB scores (domains and overall) and the adapted D-FISQ, CSDS-P, and overall GH treatment burden items. All associations were above the > 0.50 correlation: GHD-PTB Parent Emotional correlated with the adapted D-FISQ (injection domain) [0.71, $p < 0.001$], GHD-PTB Parent Emotional correlated with overall treatment burden item (0.61, $p < 0.001$), GHD-PTB Parent Interference correlated with the CSDS-P (0.63, $p < 0.001$), and GHD-PTB Parent Interference correlated with overall treatment burden item (0.61, $p < 0.001$).

Table 2 Factor analyses

	GHD-CTB–Child and GHD-CTB–Observer	Factor loadings		
		Factor 1	Factor 2	Factor 3
Emotional Well-being	CTB12 How often unhappy about how often take shots	0.808	0.163	0.024
	CTB10 How often scared about getting shots	0.794	0.103	0.263
	CTB06 How often worried about getting shots	0.748	0.123	0.301
	CTB07 How often worried about shots hurting	0.745	0.081	0.374
	CTB11 How often annoyed by shots	0.739	0.164	0.029
	CTB13 How often feel bothered different	0.486	0.421	0.015
Interference	CTB16 How often need to change plans because of shots	0.168	0.829	0.154
	CTB17 How often shots interrupt wanted to do	0.053	0.770	0.176
	CTB15 How often miss doing things because of shots	0.185	0.727	-0.092
	CTB08 How often worried about remembering to take	0.038	0.521	0.282
Physical	CTB04 How much bruising	0.186	0.097	0.775
	CTB05 How much soreness	0.153	0.316	0.768
	CTB02 How much shots sting	0.507	0.059	0.525
	CTB01 How much shots hurt	0.610	0.015	0.512
	GHD-PTB	Factor 1	Factor 2	
Emotional Well-being	PTB18 Worry causing your child pain when giving the injection	0.801	0.099	
	PTB20 Worry doing the injection correctly	0.745	0.149	
	PTB25 Often sad about your child needing injections	0.696	− 0.026	
	PTB24 Often frustrated with your child's behavior related to injections	0.642	0.083	
	PTB19 Worry remembering to give the injection	0.574	0.352	
Interference	PTB26 Often interfere with your social life	0.086	0.834	
	PTB27 Often interfere with your travel plans	0.030	0.806	
	PTB28 Often interfere with your daily routine	0.207	0.768	

CTB Child Treatment Burden, *GHD-CTB–Child* Growth Hormone Deficiency–Child Treatment Burden Measure–Child, *GHD-CTB–Observer* Growth Hormone Deficiency–Child Treatment Burden Measure–Observer, *GHD-PTB* Growth Hormone Deficiency–Parent Treatment Burden Measure, *PTB* Parent Treatment Burden

3.1.7 Known-Groups Validity

As shown in Figure S1 (see Online Resource 1 in the Electronic Supplementary Material), GHD-CTB–Child and GHD-CTB–Observer scores were able to discriminate between length of time to administer the injection (< 2 min, 2–5 min, > 5 min) in all domains and overall score ($p < 0.001$ for Physical, Emotional, and Overall, and $p < 0.01$ for Interference). The GHD-CTB–Child and GHD-CTB–Observer were not able to discriminate between age bands of when the child started GHD treatment (0.1–4.9, 5.0–8.9, and 9.0–12.4 years, $p = 0.490$ for Overall score); scores were similar across ages. While there was a trend, the GHD-CTB–Child and GHD-CTB–Observer scores were not significant by length of time on treatment (3–12, 13–23, and 24–117 months, $p = 0.389$). The Child Interference a priori hypothesis was not assessed, as 95.3% of parents/guardians reported that they did not give injections to their sleeping child.

For the GHD-PTB, the Emotional domain was able to discriminate between whether the parent/guardian gave the injections more often than the child ($p < 0.05$) [see Figure S2 in Online Resource 1]. Examining the length of time their child was on treatment (between 0 and 7 years vs 8 and 13 years), differences in GHD-PTB scores were positive (but insignificant) for the Emotional ($p = 0.441$) and Interference domains ($p = 0.852$) and the Overall score ($p = 0.490$).

3.1.8 Sensitivity to Change

Sensitivity to change was assessed by examining the change from baseline to 12 weeks for the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB follow-up scores of the treatment-naïve group.

For the GHD-CTB–Child and GHD-CTB–Observer, improvements were noted for the Overall and Emotional domains (ranging between −3.6 and −14.3 points on a 0- to 100-point scale). Associated effect sizes ranged from −0.27 to −0.57, indicating that the GHD-CTB–Child and GHD-CTB–Observer Overall and Emotional domains are sensitive to change at low to moderate levels, respectively. Small declines were seen over the 12 weeks for the Physical and Interference domains of the GHD-CTB–Child and GHD-CTB–Observer.

For the GHD-PTB, marked improvements were noted for the Emotional and Overall domains (−16.6 and −8.6 points, respectively, on a 0- to 100-point scale). Associated effect sizes were −0.74 and −0.69, indicating that the GHD-PTB is sensitive to change at high levels. The Interference domain score of the GHD-PTB had a very small improvement over 12 weeks (−0.3).

3.1.9 Interpretation of Meaningful Change

To explore MCTs, anchor-based global ratings of change were included in the study.

For the GHD-CTB–Child and GHD-CTB–Observer, using the PGIS (Table S12, see Online Resource 1 in the Electronic Supplementary Material), children and parents/guardians who indicated having a small improvement (of one or two categories) also had greater GHD-CTB–Child and GHD-CTB–Observer score improvements, except for the Interference domain (for a one-category PGIS improvement). Scores improved by 6.3 (Emotional) and 11.3 (Physical) for a one-category improvement in PGIS. Using the CGIS, improvements in GHD-CTB–Child and GHD-CTB–Observer scores were seen in the Physical and Interference domains, but not in the Emotional domain, with improvements being quite small (1.9–4.1) using the single category CGIS improvement. Changes were more robust with the PGIS versus the CGIS. GHD-CTB–Child and GHD-CTB–Observer overall score differences ranged from 3 to 7 points. Triangulating these differences (i.e., averaging between the various MCT values), we recommend an MCT be 6 points for the Overall GHD-CTB–Child and GHD-CTB–Observer scores and 6 points for the Physical domain, 9 for Emotional, and 6 for Interference.

The PGIS was also used to explore the MCT of the GHD-PTB score (Table S13, see Online Resource 1). Parents/guardians who indicated having a small improvement (of one or two categories) also had greater GHD-PTB score improvements in each domain as well as the Overall score. GHD-PTB scores improved by 4.4 points in each domain (Emotional and Interference) and Overall for a one-category improvement in PGIS. Differences ranged from 5.0 points (Interference) to 8.1 points (Emotional) for two-category improvements in PGIS. GHD-PTB overall score differences range from 4 to 14 points (examining all values in the 'Overall' column). Triangulating these differences, we recommend an MCT be 7 points for the Overall GHD-PTB score, 10 points for the Emotional domain, and 6 for Interference.

3.2 Theoretical Model and Final Validated Measure

The final validated measures include the following: the GHD-CTB–Child, a 14-item PRO for children aged 9 to < 13 years; the GHD-CTB–Observer, a 14-item ObsRO for parents/guardians of children aged 4 to < 9 years; and the GHD-PTB, an eight-item PRO measure for parents/guardians of children aged 4 to < 13 years. The theoretical model outlining the relationships between major and minor treatment burden concepts as well as consequences and modifiers to these relationships for each of these measures is presented below in Fig. 1. Additionally, the conceptual frameworks

outlining the items per domain in the final measures are shown in Fig. 2.

4 Discussion

The primary aim of this validation study was to evaluate the performance of the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB, both on an item level and at the scale level. For the GHD-CTB, the intent was for there to be only one measure, with identical items, with both a child and ObsRO version. Therefore, the items were developed to be identical, except for instructions and the perspective of who was answering the questions, for both versions. The examination of the item characteristic data supported that there were not substantial differences between what parents and children reported, which would have required examining the groups independently, and, thus, the validation was conducted on the pooled data set. As with all measures with a self-report and ObsRO version, it is always optimal for the self-report version to be used when possible. Based on concept elicitation data during the development phase of these measures, the appropriate age to begin self-report was determined to be 9 years.

The study findings demonstrated that the measures have acceptable measurement properties of item-to-item correlations, item-to-total correlations, and test–retest reliability. Although floor and ceiling effects were evidenced, all response options were used by respondents for all items. In addition, the factor analyses resulted in concordance with the original hypothesized conceptual domains and Overall score. Cronbach's alphas were all above the recommended criteria of 0.70 when examining the internal consistency of the measures.

Convergent validity was supported as the GHD-CTB−Child, GHD-CTB−Observer, and GHD-PTB scores were significantly correlated (≥ 0.40) with the DISABKIDS, CSDS, D-FISQ, and individual overall disease impact and treatment burden items. There was one correlation < 0.4 for the total score with the QoLISSY; however, it should be noted that this lower correlation of 0.342 can be considered moderate [75] and was still statistically significant as hypothesized. Further, the second hypothesized relationship for the total score with overall treatment burden had a much stronger statistically significant correlation of 0.576. Thus, we conclude that convergent validity was found to be acceptable. Known-groups validity was supported by the GHD-CTB–Observer scores that were able to discriminate between length of time to administer the injections in all domains and Overall score, and GHD-PTB scores that were able to discriminate between whether the parent/guardian gave the injections more often than the child for the Emotional domain. Non-significant trends were also found for

the GHD-CTB–Child scores by length of time on treatment, and the GHD-PTB scores for the Emotional and Interference domains, and Overall score.

Although not robust, the study did provide evidence of sensitivity to change for the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB. The brief study duration (12 weeks) and sample size may have limited the extent to which this could be observed.

Our study has developed measures to be used with children on GHD treatment aged 4 to < 13. Treatment for GHD is initiated at different ages, often depending upon which country they reside in [12] or their access to health care and health insurance. It is possible that a child younger than 4 years may be on treatment, and these measures may or may not be applicable to that younger population. Given that the age at which children start GH treatment reported in the literature for all causes of GHD is at a median age of between 5.3 and 9.7 years (depending on GHD diagnosis and sex) in the KIGS registry ($n = 83,803$) and at an average age of 9.7 years in the ANSWER and Nordinet IOS studies (combined studies $n = 37,702$) [12–14] and that the average age of treatment initiation in our study for the treatment maintenance group was 4.8 years, we feel confident that these measures will provide important information for the majority of children on treatment.

The final disease-specific GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB echo what participants reported in the concept elicitation phase of this research, which established the content validity of these instruments [44]. Together, the qualitative and psychometric studies show that the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB measures are reliable, valid measures of treatment burden for children with GHD and their parents/guardians. The growing emphasis on patient-centered outcomes, reflected by these measures, is critical for our understanding of the impact of disease on these patients and their families. Disease-specific measures allow us to better hear the patient voice and for clinicians to better understand what they can evaluate using clinical methodology. We believe that generic measures, or measures developed for other conditions, even if similar, which may also capture similar concepts, are not a substitute for a disease-specific measure if available. It should also be noted that our sample was predominately male, which is reflective of the gender distribution seen in the real world [12, 14]. Research indicates that gender ratio may be dependent on several factors, including age at treatment initiation, underlying diagnosis, country/geographic region, differences in health care systems, and gender bias. Bias towards treating boys more than girls may be due to social pressures, where height has a greater value for boys than for girls [12, 13].
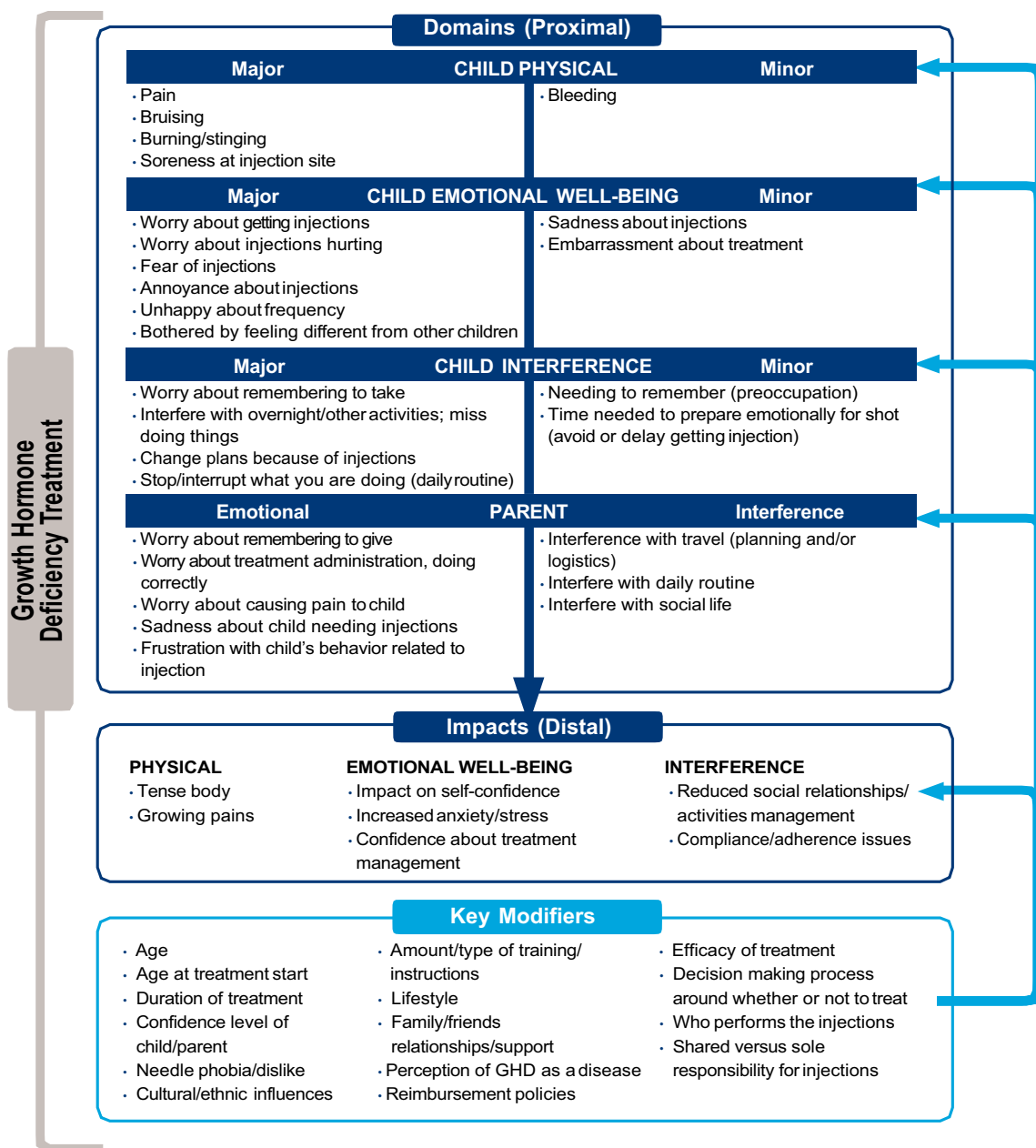
Fig. 1 Final theoretical model of GHD treatment burden (child and parent). *GHD* growth hormone deficiency

4.1 Study Limitations

Study limitations have previously been described in detail [56] and include challenges recruiting parents/guardians of children who started treatment at an early age due to US real-world clinical practice patterns, and the study population appearing to be on the healthier end of the spectrum, as suggested by 88.9% of parents who rated their child's health as either 'Excellent' or 'Very good.' Further, the sample was predominately from the US and white, which may limit the generalizability of findings. However, there were

no significant differences between the US and UK samples, although this may be due to the smaller UK sample size. Further, the concept elicitation phase included interviews with children and parents in the US, UK, and Germany, suggesting that the measures have some universality. In addition, the Interference domain score of the GHD-PTB had a very small improvement over 12 weeks, possibly due to the small proportion of parent/guardian participants in the treatment-naïve group ($n = 25$, 26%) and/or that parents learn to adapt activities that may be impacted by treatment and so avoid having the interference occur. Further, the duration

GHD-CTB	**Child Physical** *GHD-CTB – Child* *GHD-CTB – Observer*	How much shots hurt
		How much shots sting
		How much bruising
		How much soreness at injection site
	Child Emotion *GHD-CTB – Child* *GHD-CTB – Observer*	How often worried about getting shots
		How often worried about shots hurting
		How often scared about getting shots
		How often annoyed by shots
		How often unhappy about how often take shots
		How often feel bothered because different from other kids
	Child Interference *GHD-CTB – Child* *GHD-CTB – Observer*	How often worried about remembering to take
		How often miss doing things because of shots
		How often need to change plans because of shots
		How often shots interrupt/get in the way of things you wanted to do
GHD-PTB	**Parent Emotional** *GHD-PTB*	How much worry about causing child pain giving injections
		How much worry about remembering to give injection
		How much worry about doing injection correctly
		How often feel frustrated with child's behavior related to injection
		How often feel sad about child needing injections
	Parent Interference *GHD-PTB*	How often treatment interferes with your social life
		How often treatment interferes with your travel plans and vacation
		How often treatment interferes with your daily routine

Fig. 2 Conceptual frameworks of the GHD-CTB and GHD-PTB. *GHD-CTB* Growth Hormone Deficiency–Child Treatment Burden Measure, *GHD-CTB–Child* Growth Hormone Deficiency–Child Treatment Burden Measure–Child, *GHD-CTB–Observer* Growth Hormone Deficiency–Child Treatment Burden Measure–Observer, *GHD-PTB* Growth Hormone Deficiency–Parent Treatment Burden Measure

of treatment may have impacted this finding as well as the short recall period of 1 week for items such as Travel. Thus, this may not have been a suitable hypothesis for assessing improvement for the Interference domain, which may be better assessed and show greater performance when examining treatment frequency.

4.2 Clinical Implications

According to Jamie Harvey, Director of the International Coalition of Organizations Supporting Endocrine Patients (ICOSEP) and co-founder of The MAGIC Foundation, an advocacy organization that provides support services for families of children with growth disorders, when considering the patient or parent/guardian perspective and potential benefits in utilizing measures such as the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB, she responded:

'Families experience growth hormone therapy to the fullest extent. The nightly or weekly injections are carried out at home as directed by medical professionals. Physicians and nurses do not experience the child running away and hiding under beds, fights to sit still, arguments to not skipping a shot for one night, or the gut wrenching tears that some famil[ies] experience

with [...] each and every shot. They also may not be aware of children who are eager [...] for shots so that they can grow. Therefore, to facilitate the medical professional's understanding, it is helpful for them to have a form to prompt a conversation by which they can learn about the medical process occurring at home [-q]uestionnaires which can be utilized once or ongoing throughout a child's care. They offer a means to grasp and gauge each family's unique injection experience and a concrete means for helpful advice.' (J. Harvey, personal communication, January 25, 2022).

In pediatric GHD, commonly used endpoints for treatment efficacy are objective anthropometric measures of annualized height velocity and change in height SD scores. QOL metrics are equally important, but often overlooked. Currently, clinicians do not have an objective tool with which to measure treatment burden in pediatric GHD. Previously, we presented a tool for assessing the impact of treatment [56]. These clinical tools can add granularity for the clinician when evaluating a child's overall response to therapy. In addition to providing more 'real world' practical data in a clinical care setting, these tools also provide valuable QOL metrics that can be used in research. This is particularly relevant as the landscape of pediatric GHD is actively changing for the first time in over 35 years with novel long-acting GH therapies in development. By using both traditional anthropometric measures in conjunction with QOL metrics, researchers can better characterize the true impact of these emerging therapies.

4.3 Future Research

Understanding the patient perspective is an iterative process. This study has taken the first step in understanding treatment burden in children diagnosed with GHD and their parents and raises several important questions that should be explored in future research. First, as any response to treatment is influenced by cultural as well as clinical factors, we suggest that additional cross-cultural studies, including non-white and non-Western populations, would be informative and help our understanding of which treatment burden aspects are more cultural and what are more disease driven. Further, studies of a longer duration and with larger and more varied populations of treatment-naïve patients and greater diversity of health status may eliminate some of the issues around floor and ceiling effects as well as help improve our understanding of the responsiveness of these measures to changes due to treatment initiation over time. Additionally, our research was based on children up to age 13; however, children may stay on GH treatment through adolescence as continued growth remains possible and little is known about treatment burden in this age group. The

applicability of these new child and parent treatment burden measures for older children on GH treatment should be explored.

5 Conclusion

The cumulative evidence on the psychometric properties of the GHD-CTB–Child, GHD-CTB–Observer, and GHD-PTB support the validity of their use as PRO and ObsRO measures to capture the experiences associated with the treatment burden for children with GHD and their parents/guardians in both clinical and research settings. Better health-related QOL assessment of treatment burden will also allow clinicians to have targeted discussions with their patients regarding their experience with treatment, which should improve provider–patient communications as well as improve adherence to treatment and treatment outcomes. Disease-specific assessment of treatment burden in clinical trials may provide data suggesting which treatments provide a better patient experience as well as improved growth.

Hospital for Children, London; Jeremy Kirk, MD FRCP FRCPCH DCH-Birmingham Women's and Children's Hospital, Birmingham.

Declarations

Conflict of interest M. Brod, S. Alolga, J.F. Beck, and D.M. Bushnell are paid consultants to the pharmaceutical industry, including Novo Nordisk. M. Højby Rasmussen is an employee of Novo Nordisk A/S and owns stock in the company. K.W. Lee was an employee of Novo Nordisk A/S when the research was conducted. A. 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 #: 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 the publication may be available on a case-by-case basis, for reasonable requests, from the corresponding author.

Code availability Not applicable.

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. Brod M, Alolga SL, Beck JF, Wilkinson L, Hojbjerre 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.
3. Stabler B, Clopper RR, Siegel PT, Stoppani C, Compton PG, Underwood LE. Academic achievement and psychological adjustment in short children. The National Cooperative Growth Study. J Dev Behav Pediatr. 1994;15(1):1–6. https://doi.org/10.1097/00004703-199402000-00001.
4. Butler G, Turlejski T, Wales G, Bailey L, Wright N. Growth hormone treatment and health-related quality of life in children and adolescents: a national, prospective, one-year controlled study. Clin Endocrinol. 2019;91(2):304–13. https://doi.org/10.1111/cen.14011.
5. Geisler A, Lass N, Reinsch N, Uysal Y, Singer V, Ravens-Sieberer U, et al. Quality of life in children and adolescents with growth hormone deficiency: association with growth hormone treatment. Horm Res Paediatr. 2012;78(2):94–9. https://doi.org/10.1159/000341151.
6. Sommer R, Bullinger M, Chaplin J, Do JK, Power M, Pleil A, et al. Experiencing health-related quality of life in paediatric short stature - a cross-cultural analysis of statements from patients and parents. Clin Psychol Psychother. 2017;24(6):1370–6. https://doi.org/10.1002/cpp.2105.
7. Stagi S, Scalini P, Farello G, Verrotti A. Possible effects of an early diagnosis and treatment in patients with growth hormone deficiency: the state of art. Ital J Pediatr. 2017;43(1):81. https://doi.org/10.1186/s13052-017-0402-8.
8. Ciresi A, Amato MC, Criscimanna A, Mattina A, Vetro C, Galluzzo A, et al. Metabolic parameters and adipokine profile during GH replacement therapy in children with GH deficiency. Eur J Endocrinol. 2007;156(3):353–60. https://doi.org/10.1530/eje.1.02343.
9. 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.
10. 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.
11. 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.
12. Ranke MB, Lindberg A, Tanaka T, Camacho-Hubner C, Dunger DB, Geffner ME. Baseline characteristics and gender differences in prepubertal children treated with growth hormone in Europe, USA, and Japan: 25 Years' KIGS(R) experience (1987–2012) and review. Horm Res Paediatr. 2017;87(1):30–41. https://doi.org/10.1159/000452887.
13. 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;93(6):2050–6. https://doi.org/10.1210/jc.2007-2617.
14. Savendahl L, Polak M, Backeljauw P, Blair JC, Miller BS, Rohrer TR, et al. Long-term safety of growth hormone treatment in childhood: two large observational studies: NordiNet IOS and ANSWER. J Clin Endocrinol Metab. 2021;106(6):1728–41. https://doi.org/10.1210/clinem/dgab080.
15. Laron Z. Growth hormone therapy: emerging dilemmas. Pediatric Endocrinol Rev. 2011;8(4):364–73.
16. Collett-Solberg PF, Ambler G, Backeljauw PF, Bidlingmaier M, Biller BMK, Boguszewski MCS, 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. https://doi.org/10.1159/000502231.

17. 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.

18. Stochholm K, Kiess W. Long-term safety of growth hormone-A combined registry analysis. Clin Endocrinol. 2018;88(4):515–28. https://doi.org/10.1111/cen.13502.

19. Child CJ, Zimmermann AG, Chrousos GP, Cummings E, Deal CL, Hasegawa T, et al. Safety outcomes during pediatric GH therapy: final results from the prospective GeNeSIS observational program. J Clin Endocrinol Metab. 2019;104(2):379–89. https://doi.org/10.1210/jc.2018-01189.

20. 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;174(2):P1-9. https://doi.org/10.1530/eje-15-0873.

21. 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.

22. Richmond E, Rogol AD. Treatment of growth hormone deficiency in children, adolescents and at the transitional age. Best Pract Res Clin Endocrinol Metab. 2016;30(6):749–55. https://doi.org/10.1016/j.beem.2016.11.005.

23. Abe S, Okumura A, Mukae T, Nakazawa T, Niijima S, Yamashiro Y, et al. Depressive tendency in children with growth hormone deficiency. J Paediatr Child Health. 2009;45(11):636–40. https://doi.org/10.1111/j.1440-1754.2009.01586.x.

24. Chaplin JE, Kriström B, Jonsson B, Hägglöf B, Tuvemo T, Aronson AS, et al. Improvements in behaviour and self-esteem following growth hormone treatment in short prepubertal children. Horm Res Paediatr. 2011;75(4):291–303. https://doi.org/10.1159/000322937.

25. Hodax JK, DiVall SA. Update on methods to enhance growth. Curr Opin Endocrinol Diabetes Obes. 2020;27(1):82–6. https://doi.org/10.1097/med.0000000000000513.

26. Kaplowitz P, Manjelievskaia J, Lopez-Gonzalez L, Morrow CD, Pitukcheewanont P, Smith A. Economic burden of growth hormone deficiency in a US pediatric population. J Manag Care Spec Pharm. 2021;27(8):1118–28. https://doi.org/10.18553/jmcp.2021.21030.

27. Acerini CL, Segal D, Criseno S, Takasawa K, Nedjatian N, Rohrich S, et al. Shared decision-making in growth hormone therapy-implications for patient care. Front Endocrinol (Lausanne). 2018;9:688. https://doi.org/10.3389/fendo.2018.00688.

28. Yuen KCJ, Miller BS, Biller BMK. The current state of long-acting growth hormone preparations for growth hormone therapy. Curr Opin Endocrinol Diabetes Obes. 2018;25(4):267–73. https://doi.org/10.1097/MED.0000000000000416.

29. Aydın BK, Aycan Z, Sıklar Z, Berberoğlu M, Ocal G, Cetinkaya S, et al. Adherence to growth hormone therapy: results of a multicenter study. Endocr Pract. 2014;20(1):46–51. https://doi.org/10.4158/ep13194.or.

30. Desrosiers P, O'Brien F, Blethen S. Patient outcomes in the GHMonitor: the effect of delivery device on compliance and growth. Pediatric Endocrinol Rev. 2005;2(Suppl 3):327–31.

31. Blanco-López A, Antillón-Ferreira C, Saavedra-Castillo E, Barrientos-Pérez M, Rivero-Escalante H, Flores-Caloca O, et al. Adherence to treatment in children with growth hormone deficiency, small for gestational age and Turner syndrome in Mexico: results of the Easypod™ connect observational study (ECOS). J Endocrinol Invest. 2020;43(10):1447–52. https://doi.org/10.1007/s40618-020-01218-4.

32. Mohseni S, Heydari Z, Qorbani M, Radfar M. Adherence to growth hormone therapy in children and its potential barriers. J Pediatr Endocrinol Metabol. 2018;31(1):13–20. https://doi.org/10.1515/jpem-2017-0157.

33. Kapoor RR, Burke SA, Sparrow SE, Hughes IA, Dunger DB, Ong KK, et al. Monitoring of concordance in growth hormone therapy. Arch Dis Child. 2008;93(2):147–8. https://doi.org/10.1136/adc.2006.114249.

34. Rosenfeld RG, Bakker B. Compliance and persistence in pediatric and adult patients receiving growth hormone therapy. Endocr Pract. 2008;14(2):143–54. https://doi.org/10.4158/ep.14.2.143.

35. Wit JM, Deeb A, Bin-Abbas B, Al Mutair A, Koledova E, Savage MO. Achieving optimal short- and long-term responses to paediatric growth hormone therapy. J Clin Res Pediatr Endocrinol. 2019;11(4):329–40. https://doi.org/10.4274/jcrpe.galenos.2019.2019.0088.

36. Haverkamp F, Gasteyger C. A review of biopsychosocial strategies to prevent and overcome early-recognized poor adherence in growth hormone therapy of children. J Med Econ. 2011;14(4):448–57. https://doi.org/10.3111/13696998.2011.590829.

37. Fisher BG, Acerini CL. Understanding the growth hormone therapy adherence paradigm: a systematic review. Horm Res Paediatr. 2013;79(4):189–96. https://doi.org/10.1159/000350251.

38. Norgren S. Adherence remains a challenge for patients receiving growth hormone therapy. Pediatr Endocrinol Rev. 2009;6(Suppl 4):545–8.

39. Lal RA, Hoffman AR. Perspectives on long-acting growth hormone therapy in children and adults. Arch Endocrinol Metabol. 2019;63(6):601–7. https://doi.org/10.20945/2359-3997000000190.

40. Haverkamp F, Johansson L, Dumas H, Langham S, Tauber M, Veimo D, et al. Observations of nonadherence to recombinant human growth hormone therapy in clinical practice. Clin Ther. 2008;30(2):307–16. https://doi.org/10.1016/j.clinthera.2008.02.017.

41. Gau M, Takasawa K. Initial patient choice of a growth hormone device improves child and adolescent adherence to and therapeutic effects of growth hormone replacement therapy. J Pediatr Endocrinol Metabol. 2017;30(9):989–93. https://doi.org/10.1515/jpem-2017-0146.

42. Reiter EO, Attie KM, Moshang T Jr, Silverman BL, Kemp SF, Neuwirth RB, et al. A multicenter study of the efficacy and safety of sustained release GH in the treatment of naive pediatric patients with GH deficiency. J Clin Endocrinol Metab. 2001;86(10):4700–6. https://doi.org/10.1210/jcem.86.10.7932.

43. Gonzalez Briceno LG, Viaud M, Beltrand J, Flechtner I, Dassa Y, Samara-Boustani D, et al. Improved general and height-specific quality of life in children with short stature after 1 year on growth hormone. J Clin Endocrinol Metab. 2019;104(6):2103–11. https://doi.org/10.1210/jc.2018-02523.

44. Brod M, Højbjerre L, Alolga SL, Beck JF, Wilkinson L, Rasmussen MH. Understanding treatment burden for children treated for growth hormone deficiency. The Patient. 2017;10(5):653–66. https://doi.org/10.1007/s40271-017-0237-9.

45. Carcone AI, Ellis DA, Naar-King S. Linking caregiver strain to diabetes illness management and health outcomes in a sample of adolescents in chronically poor metabolic control. J Dev Behav Pediatr. 2012;33(4):343–51. https://doi.org/10.1097/DBP.0b013e31824eaac8.

46. Young MT, Lord JH, Patel NJ, Gruhn MA, Jaser SS. Good cop, bad cop: quality of parental involvement in type 1 diabetes management in youth. Curr Diab Rep. 2014;14(11):546. https://doi.org/10.1007/s11892-014-0546-5.

47. Wiebe DJ, Chow CM, Palmer DL, Butner J, Butler JM, Osborn P, et al. Developmental processes associated with longitudinal

declines in parental responsibility and adherence to type 1 diabetes management across adolescence. J Pediatr Psychol. 2014;39(5):532–41. https://doi.org/10.1093/jpepsy/jsu006.

48. Brod M, Perwien A, Adler L, Spencer T, Johnston J. Conceptualization and assessment of quality of life for adults with attention-deficit/hyperactivity disorder. Prim Psychiatry. 2005;2:58–64.

49. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. Med Care. 1989;27(3 Suppl):217–32. https://doi.org/10.1097/00005650-198903001-00018.

50. US Food and Drug Administration. 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.

51. Savendahl L, Polak M, Backeljauw P, Blair J, Miller BS, Rohrer TR, et al. Treatment of children with GH in the United States and Europe: long-term follow-up from NordiNet(R) IOS and ANSWER program. J Clin Endocrinol Metab. 2019;104(10):4730–42. https://doi.org/10.1210/jc.2019-00775.

52. Matza LS, Patrick DL, Riley AW, Alexander JJ, Rajmil L, Pleil AM, Bullinger M. 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.

53. 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. Accessed 20 Feb 2019.

54. 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.

55. U.S. Office of the Federal Register. Electronic Code of Federal Regulations. Titles 21§50 Protection of human subjects and §56 Institutional review boards. Office of the Federal Register, Washington, D.C. 2018. https://www.ecfr.gov/cgi-bin/text-idx?SID=25670b7f62f18a7a8771e34b0ece4cbe&mc=true&tpl=/ecfrbrowse/Title21/21cfrv1_02.tpl#0. Accessed 10 May 2019.

56. Brod M, Højby Rasmussen M, Vad K, Alolga S, Bushnell DM, Bedoin J, et al. Psychometric validation of the growth hormone deficiency-child impact measure (GHD-CIM). PharmacoEconomics Open. 2021;5(3):505–18. https://doi.org/10.1007/s41669-020-00252-5.

57. Bentler PM, Chou CP. Practical issues in structural modeling. Sociol Methods Res. 1987;16:78. https://doi.org/10.1177/0049124187016001004.

58. Linacre JM. Sample size and item calibration stability. Rasch Measurement Trans. 1994;7:4.

59. 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.

60. 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.

61. 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.

62. 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.

63. SPSS Inc. PASW statistics for windows. 18.0 ed. Chicago: SPSS Inc.; 2009.

64. Carmines EC, Zeller RA. Reliability and validity assessment. Beverly Hills: Sage Publications; 1979.

65. 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. p. 82–8.

66. Byrne BM. Structural equation modeling with AMOS: basic concepts, applications, and programming. New York: Routledge; 2010.

67. Arbuckle JL. AMOS 20.0 user's guide. Armonk, N.Y.: IBM Corporation; 2011.

68. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16(2):297–334.

69. Hays RD, Revicki DA. Reliability and validity (including responsiveness). In: Fayers PM, Hays RD, editors. Assessing quality of life in clinical trials. New York: Oxford; 2005. p. 25–39.

70. Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, de Vet HCW et al. COSMIN methodology for systematic reviews of Patient-reported Outcome Measures (PROMs): user manual. 2018.

71. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting G. 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.

72. 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/10.1097/00005650-19990 5000-00006.

73. Kline RB. Principles and practice of structural equation modeling. 3rd ed. New York: Guilford Publications; 2010.

74. 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.

75. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale: L. Erlbaum Associates; 1988.

Authors and Affiliations

Meryl Brod[1] · **Michael Højby Rasmussen[2]** · **Suzanne Alolga[1]** · **Jane F. Beck[1]** · **Donald M. Bushnell[3]** · **Kai Wai Lee[2,5]** · **Aristides Maniatis[4]**

✉ Meryl Brod
  mbrod@thebrodgroup.net

[1] The Brod Group, Mill Valley, CA, USA

[2] Novo Nordisk A/S, Søborg, Denmark

[3] Evidera | PPD, Bethesda, MD, USA

[4] Rocky Mountain Pediatric Endocrinology, Centennial, CO, USA

[5] Present Address: Biopharmaceuticals R&D, AstraZeneca, Cambridge, UK