

Assessing the Impact of Growth Hormone Deficiency (GHD) in Adults: Interpreting Change of the Treatment-Related Impact Measure—Adult Growth Hormone Deficiency (TRIM-AGHD)

Meryl Brod¹  · Jane F. Beck¹ · Lise Højbjerg² · Donald M. Bushnell³ · Johan Erpur Adalsteinsson² · Lars Wilkinson² · Michael Højby Rasmussen²

Published online: 23 May 2018
© The Author(s) 2018

Abstract

Background This study's purpose was to assess the minimal important difference (MID) for the Treatment-Related Impact Measure—Adult Growth Hormone Deficiency (TRIM-AGHD), a patient-reported outcome measure assessing growth hormone deficiency (GHD) impacts. The measure was demonstrated to have adequate psychometric measurement properties, and be reliable and valid. For scores to be interpretable, the TRIM-AGHD must be responsive to treatment benefit and the MID in scores quantified.

Methods A prospective, non-interventional, observational, clinic-based survey study of naïve-to-treatment adult GHD patients ($N = 98$) was conducted. Key assessments were at baseline and follow-up (between 4 and approximately 8 weeks), with weekly telephone monitoring post-baseline (last $n = 34$ patients). Responsiveness was evaluated using the effect size of change scores from baseline to follow-up. MID estimates were derived from distribution-based (half standard deviation [0.5 SD], standard error of measurement [SEm]) and anchor-based methods (patient global rating of change [PGRC]) using change scores from baseline to initial report of minimal improvement in GHD severity.

Findings from each method were converged to establish an acceptable MID.

Results Patients were mean age 49.7 years, 65.6% female, and 76.0% Caucasian. The TRIM-AGHD was highly responsive to treatment with the total score effect size being 1.38. For the total score, the 0.5 SD was 8.09 and the SEm was 2.66. The difference found using the PGRC was 20.43. The converged MID value for the total score was 10 points.

Conclusions The TRIM-AGHD is a highly responsive measure assessing AGHD treatment impacts. A 10-point change score is considered a clinically meaningful improvement.

Johan Erpur Adalsteinsson at Novo Nordisk A/S at the time of the study.

✉ Meryl Brod
mbrod@thebrodgroup.net

¹ The Brod Group, 219 Julia Avenue, Mill Valley, CA 94941, USA

² Novo Nordisk A/S, Vandtaarnsvej 114, 2860 Soeborg, Denmark

³ Health Research Associates, Inc, 6505 216th Street SW, Suite 105, Mountlake Terrace, WA 98043, USA

Key Points for Decision Makers

For survey scores to be interpretable, the minimal important difference (MID) in scores must be quantified, which is the smallest change in a score for a patient that indicates an actual change between two time points.

The MID allows both researchers and clinicians to interpret if treatment plans are effective.

The Treatment-Related Impact Measure—Adult Growth Hormone Deficiency MID has been determined to be an improvement of 10 points in the total score.

1 Background

Adult growth hormone deficiency (AGHD) is a debilitating, rare disease with the incidence of genuine adult-onset growth hormone deficiency (GHD) estimated at 10 per million [1]. Over 50,000 adults in the United States are growth hormone deficient, with approximately 6000 new cases reported yearly (this figure includes children with growth hormone deficiency that are transitioning to adulthood) [2].

AGHD is associated with reduced muscle mass and muscle strength, reduced bone mass or osteoporosis, and an increase in body fat [3, 4]. Additionally, AGHD is associated with increased cardiovascular morbidity and mortality [5]. Findings regarding the role of height as a predictor of health-related quality of life (HRQoL) in AGHD are inconclusive [6, 7], suggesting that other factors may be more central to understanding HRQoL in this population. AGHD is also associated with impaired concentration and loss of memory, dissatisfaction with body image, and decreased quality of life (QoL) [8]. Important areas of impact for AGHD include energy or vitality levels, mood, social isolation, and self-control [9, 10]. Adults with AGHD may also experience psychological impairments such as depression, anxiety, and social isolation [11–14]. Negative impacts of AGHD on patient physiology, psychological well-being, cognitive functioning, and QoL are well documented; however, studies suggest that growth hormone (GH) treatments can be effective in mitigating poorer health outcomes [15, 16]. In fact, a recent study investigated the long-term effects of growth hormone treatment on adults with GHD and found the effect of GH therapy on QoL showed sequential improvements and stabilization until 6-year follow-up [17].

The Treatment-Related Impact Measure–Adult Growth Hormone Deficiency (TRIM-AGHD), a patient-reported outcome (PRO) measure of the impact of GHD on adult functioning and wellbeing has been previously developed according to the US Food and Drug Administration (FDA) guidelines for the development of PRO measures for use in clinical trials and labelling [18] and found to be conceptually sound and psychometrically reliable and valid [19]. Beginning with qualitative interviews to elicit concepts, then utilizing quantitative data for item reduction and defining the measurement model, the TRIM-AGHD was developed. It is a 27-item measure with four domains (Energy, Psychological, Cognitive, and Physical). It is primarily scored independently for each domain with score ranges of 0–100 (lower scores indicating a better health state). As evidenced with results from a higher-order factor analysis, a total score (range 0–100) can also be reported [19].

As an additional step in the validation of the TRIM-AGHD, this study assessed the sensitivity to change (responsiveness) and minimal important difference (MID) for the measure. The assessment of these psychometric properties for a newly developed PRO is an important scientific practice to ensure meaningfulness and interpretability of the measure [20, 21]. Without this information, it would be problematic for both clinicians and researchers to interpret if treatment plans are effective and if so, if that effectiveness is meaningful to patients. MID can be estimated using both anchor- and distribution-based approaches [22]. The FDA recommends that distribution-based methods for determining clinical significance of changes in scores should be considered as supportive and not the only basis for determining a responder definition [18]. Anchor-based methods assess responsiveness in relation to an independent measure (e.g., external rating) to quantify the meaning of a particular degree of change in the health construct [23–26]. Distribution-based methods rely on the distribution of scores within a population and relate clinical significance to a change in magnitude at least equal to a statistical parameter of group data such as variability (e.g., standard deviation) or reliability (e.g., Cronbach's α). When each of these approaches are used to determine the MID of a scale, a range of values rather than a single point estimate are expected [25]. However, a more narrow range or even a single point of MID would be more helpful than the broad range determined by multiple estimates. A triangulation-type method, using anchor-based and distribution-based (both half a standard deviation and standard error of measurement [SEm]), can be used to converge on a reasonable MID [27].

The purpose of this study was to calculate an MID using a triangulation method and to examine responsiveness of the TRIM-AGHD. The primary hypothesis for this study was that as evaluations of severity improve, for example, on one of the measures utilized in this study—the Patient Global Rating of Change (PGRC), treatment-related impact (assessed by the TRIM-AGHD) will improve.

2 Methods

This was a prospective, non-interventional, observational, clinic-based survey study of GHD patients who were starting a new treatment for their GHD at the time of enrolment. Enrolments and study assessments were conducted between March 2014 and December 2015. Eligible patients were recruited by physicians at four study sites (one academic and three private-practice settings) in the US (Los Angeles, CA; Oklahoma City, OK; Salt Lake City, UT; and Dearborn, MI) from their current patient caseload or identified by chart review. Patients were invited to

participate by designated site personnel who had completed telephone-based training to review the study protocol and review recruitment procedures. The decision to initiate treatment, and the choice of treatment, was made by the physician as per usual care and independent of the patient's decision to participate in the study. In order not to influence treatment choice, the patient was asked to participate in the study only after the decision to treat his/her GHD and the specific treatment had been agreed upon by the patient and the physician.

Patients were included in the study if they were a male or female of at least 23 years of age and not more than 79 years of age; were able to speak, read, and write in English; had either adult or childhood onset of GHD and a confirmed GHD diagnosis; were GHD treatment naïve (not being on a prescription treatment for their GHD currently and for at least 6 months); were beginning a new prescription GHD treatment and expected to be on this treatment for GHD for a minimum of 6 months; and completed the informed consent before any study-related activities. Patients were excluded if they were on a prescription medication for treatment of GHD in past 6 months; had a Beck Depression Inventory II (BDI-II) score > 25 at enrolment; were female and pregnant or intending to become pregnant or were breastfeeding or not using adequate contraceptive methods; had acute severe illness associated with weight loss in the last 6 months (defined as a loss of more than 5.0% total body weight); had active Cushing's syndrome within the last 24 months; had overt diabetes mellitus; had a mental incapacity, unwillingness or language barrier precluding adequate understanding or cooperation; or had previously participated in this study.

At the in-person, baseline visit, all eligible and interested patients signed the informed consent and were then enrolled after completing the screening process. As part of the screening process, all patients had their diagnosis of GHD and all eligibility criteria confirmed by their physician, including the BDI-II. Enrolled patients completed the baseline questionnaire battery, which included the TRIM-AGHD, one-item Patient Global Impression (PGI) of severity, and a brief demographic form. The clinician completed a Clinician Global Impression (CGI) of severity and a brief medical information sheet, which included physical measurements (e.g., blinded waist circumference) and details on prescribed GHD medication.

At the follow-up visit, the patients completed the follow-up questionnaire, which included the TRIM-AGHD, the PGI, and the PGRC. The PGRC has 15 response options ranging from 'a very great deal better' to 'a very great deal worse' with 'no change' in the middle. In this analysis, minimal improvement was defined as patients indicating 'Almost the same/hardly better at all', 'A little

better', or 'Somewhat better' [28]. The clinician completed the CGI and assessed a blinded waist circumference.

All patients were expected to complete a follow-up visit at week 8 to assess the MID. After the first six patients were assessed at week 8, the schedule of assessments was modified due to reports from two patients of more than minimal improvement in the anchor questionnaire (PGRC). Subsequently, assessments occurred between approximately weeks 4–8 in order to detect the earliest minimal improvement.

Since more than minimal improvement on the PGRC questionnaire was reported at week 4 ± 1 day by seven patients (among the first 62 patients assessed), it was determined that to best capture the earliest time point where minimal improvement was occurring for the last 34 patients, the assessment strategy had to be altered again. The last 34 patients were monitored weekly by telephone starting 1 week after the baseline visit. During these calls, which were conducted by study-site personnel, all patients completed the PGRC. If patients reported minimal improvement in their disease status between weeks 2 and 7, they were brought back into the clinic for the follow-up visit as soon as possible, within 5 business days, to further evaluate the PGI, CGI, and TRIM-AGHD. For this subset of participants, if no minimal improvement was reported by week 7, their follow-up assessment was completed at week 8 ± 1 week.

Patients were treated as per usual care with no intervention by the study. This study was conducted in accordance with the Declaration of Helsinki and the Guidelines for Good Pharmacoepidemiology Practices [29, 30] and was approved by Copernicus Group IRB (approval TBG1-13-475). In addition, the COSMIN checklist (CONsensus-based Standards for the selection of health Measurement INstruments) [31, 32] was reviewed for design requirements in the assessment of responsiveness.

2.1 Statistical Methods

Sensitivity to change, the ability of an instrument to detect small but important changes, was evaluated using the effect size (ES) [33]. For this index, the numerator was the mean baseline to endpoint change and the denominator was the standard deviation (SD) at baseline. Higher values for the ES indicated a greater sensitivity to change. For the ES, Cohen [34] provided guidance on interpretation of the magnitude, where a 0.20 ES was considered a small change, 0.50 a moderate change, and 0.80 a large change.

The MID of the TRIM-AGHD was assessed using both anchor-based and distribution-based techniques [22, 23, 25]. Distribution-based methods included (i) the examination of 0.5 SD of the change between assessments [35] and (ii) the SEM, which is the observed SD multiplied

by the square root of 1 minus the reliability (where reliability is represented by the Cronbach's alpha coefficient) [36, 37]. For the anchor-based calculation, the MID was assessed using reports from the PGRC, which should represent perceptions of change due to treatment. Minimally important changes were calculated between the initial baseline assessment and the follow-up assessment at the time each patient registered an improvement in their disease status (up to 8 weeks). Change was inherent in the PGRC item, with patients indicating whether their GHD condition had stayed about the same, gotten better, or worsened. The anchor-based MID was calculated as the difference between the improved group ('Almost the same/hardly better at all', 'A little better', and 'Somewhat better') and the group who stayed the same. Given that not one of these methods are more psychometrically robust than another, and in an effort to arrive at a single estimate, we examined each of the different MIDs found using the various methods, averaged them, and rounded to the nearest integer. Therefore, by triangulating, this approach takes into consideration what would be considered both clinically meaningful and perceived as beneficial from the patient's viewpoint.

3 Results

3.1 Sample Characteristics

Ninety-eight patients were confirmed eligible and enrolled in the study. Data for 96 patients with post-baseline values comprise the full analysis set used for the MID determination. These patients completed questionnaires including the TRIM-AGHD at baseline and at follow-up between 4 and approximately 8 weeks, and 247 monitoring telephone calls were completed between the baseline and follow-up visits. The average time between baseline (treatment initiation) and follow-up visit was 6.57 weeks. As seen in Table 1, mean age was 49.7 years (range 29–68) with 65.6% being female and 76.0% being Caucasian. Global impression of severity at baseline was rated as 'Very severe' by 85.4% of patients and 89.6% of their clinicians. The primary cause of GHD was idiopathic in nature (46.9%). Two recruited individuals were withdrawn from the study (one was lost to follow-up and one discontinued medication).

3.2 Responsiveness and MID

At follow-up, the TRIM-AGHD was shown to be highly responsive ($ES > 0.80$) to treatment with the total score effect size being 1.38 (subscales ranged between 1.22 and 1.36, see Table 2). For distribution-based MID

calculations, 0.5 SD and SEM were examined. As shown in Table 2, the 0.5 SD for the TRIM-AGHD total score was 8.09 (subscales ranged between 8.44 and 9.18). The SEM for the TRIM-AGHD total score was 2.66 (subscales ranged between 3.55 and 4.57). Examining the anchor-based method using the PGRC, differences were larger. The majority of patients indicated getting 'Better' (59.4%) or were 'About the same' (34.4%). Only six patients (6.3%) indicated a worsening of their GHD. As shown in Table 3, the difference in TRIM-AGHD total score between the 'Better' group and the 'About the same' group was 20.43 (subscales ranged between 19.63 and 21.80).

3.3 MID Estimate

Averaging each of the MIDs from all approaches (SEM, 0.5 SD, and the anchor-based PGRC), the MID based on the convergence of these values was for each of the subscales: Energy (11.45), Psychological (10.65), Cognitive (11.28), and Physical (11.11), and for the total score (10.40). Rounding to the nearest integer for these values suggests that the MID value for each of the subscales (Energy, Psychology, Cognitive, and Physical) should be 11 points and the MID value of 10 points should be used for the total score. These values are believed to be clinically meaningful and would be perceived as beneficial from the patient's viewpoint.

3.4 Waist Circumference

TRIM-AGHD scores were also evaluated in relation to changes in waist circumference (patients who either increased in waist circumference [between 0.50 and 3.25 inches], decreased [between 0.40 and 10.00 inches], or remained with the same circumference). As seen in Fig. 1, TRIM-AGHD scores were sensitive to waist circumference change. All patients had improvements in TRIM-AGHD scores, but patients who had a decrease in waist circumference had statistically significant improvements ($p < 0.01$) compared with patients who had an increase in waist circumference.

4 Discussion

As the number of clinical trials in AGHD increases, determining the MID of instruments used to measure GHD response is prudent for the conduct and interpretability of meaningful future clinical trials. Additionally, it is important for clinicians who treat these patients to assess treatment benefit over the course of treatment so that targeted treatment strategies can be implemented. For adults with GHD, improvement in height is not an appropriate

Table 1 Demographic and clinical characteristics

	Total (N = 96) ^a
Age ^b , mean (range)	49.7 (29–68)
Age when diagnosed ^c , mean (range)	48.5 (9–68)
Gender, # (%)	
Female	63 (65.6)
Male	33 (34.4)
Marital status, # (%)	
Single	46 (47.9)
Married/partnered	45 (46.9)
Divorced	5 (5.2)
Ethnicity, # (%)	
Caucasian	73 (76.0)
Black/African American	13 (13.5)
Latino/Hispanic/Mexican–American	4 (4.2)
Asian/Pacific Islander	2 (2.1)
Native American/Alaskan native	3 (3.1)
Declined to answer	1 (1.0)
Work status, # (%)	
Full-time paid position	81 (84.4)
Part-time paid position	4 (4.2)
Not currently working for pay	11 (11.5)
Education, # (%)	
Grade school or less	8 (8.3)
High school or technical school	13 (13.5)
College	54 (56.3)
Graduate or professional school	19 (19.8)
Blank response	2 (2.1)
Income, # (%)	
Less than \$20,000	3 (3.1)
\$20,000–\$39,999	2 (2.1)
\$40,000–\$59,999	14 (14.6)
\$60,000–\$79,999	7 (7.3)
\$80,000–\$99,999	9 (9.4)
\$100,000+	5 (5.2)
Declined to answer/blank response	56 (58.3)
Beck Depression Inventory (BDI-II), mean (range)	16.0 (4–25)
Patient Global Impression of Severity at baseline, # (%)	
Moderate	5 (5.2)
Severe	8 (8.3)
Very severe	82 (85.4)
Missing	1 (1.0)
Clinician Global Impression of Severity at baseline, # (%)	
Moderate	5 (5.2)
Severe	5 (5.2)
Very severe	86 (89.6)
Cause of GHD, # (%): patient reported	
Unknown cause	45 (46.9)
Pituitary tumor	9 (9.4)
Other growths/disorders in the pituitary or hypothalamic area	8 (8.3)

Table 1 continued

	Total (N = 96) ^a
History of surgery or radiation therapy to the pituitary area or base of skull	3 (3.1)
Head trauma	17 (17.7)
Other	13 (13.5)
Blank response	1 (1.0)
Medical conditions (not including GHD), # (%) ^d	
Arthritis, rheumatic diseases, musculoskeletal conditions	11 (11.5)
Cancer	2 (2.1)
Heart disease, cardiovascular conditions (including hypertension)	52 (54.2)
Ear, nose and throat conditions	13 (13.5)
Endocrine disorders (including diabetes and thyroid disorders)	43 (44.8)
Eye disorders	44 (45.8)
Internal organ disease (liver, pancreas, spleen, etc.)	1 (1.0)
Kidney disease, urinary conditions	3 (3.1)
Lung disease, respiratory conditions (including allergies and asthma)	41 (42.7)
Mental health conditions (including depression and anxiety)	6 (6.3)
Metabolic conditions (including elevated cholesterol)	37 (38.5)
Stomach, intestinal, gastrointestinal disease	22 (22.9)
Stroke, neurological condition	0 (0.0)
Other condition	3 (3.1)
None of these	13 (13.5)
No. of medical conditions (not including GHD), mean (range)	2.9 (0–9)
No. of prescription medications, besides GHD prescription, regularly taken in past 3 months ^b ; mean (range)	3.1 (0–19)

BDI-II Beck Depression Inventory–II; GHD growth hormone deficiency

^aN = 96 comprises the full analysis set, and not the enrolled population; recruitment totals for completing subjects by site: two sites n = 1 each, one site n = 5, one site n = 89

^bN = 95 due to blank response

^cN = 94 due to two blank responses

^dResponses not mutually exclusive except ‘none of these’, which was an exclusive response

Table 2 TRIM-AGHD: sensitivity to change

TRIM-AGHD subscale	Assessment 1 (baseline)		Assessment 2		Change (2–1)		Cronbach's alpha	Effect size ^a	Distribution-based MID	
	Mean	SD	Mean	SD	Mean	SD			0.5 SD	SEm ^b
Energy	85.3	16.2	63.4	18.0	– 21.9	18.0	0.96	1.22	9.00	3.55
Psychological	81.1	15.5	59.5	15.3	– 21.6	16.9	0.95	1.28	8.44	3.89
Cognitive	86.6	17.7	63.1	17.6	– 23.5	18.4	0.96	1.28	9.18	3.85
Physical	88.3	16.9	65.1	16.9	– 23.2	17.1	0.93	1.36	8.57	4.57
Total score	84.2	14.5	61.9	15.4	– 22.3	16.2	0.97	1.38	8.09	2.66

MID minimal important difference, SD standard deviation, SEM standard error of measurement, TRIM-AGHD Treatment Related Impact Measure—Adult Growth Hormone Deficiency

^aEffect size = mean change divided by the standard deviation of Assessment 1

^bSEM = observed SD × Sqrt(1 – reliability) where the reliability is represented by the Cronbach's alpha

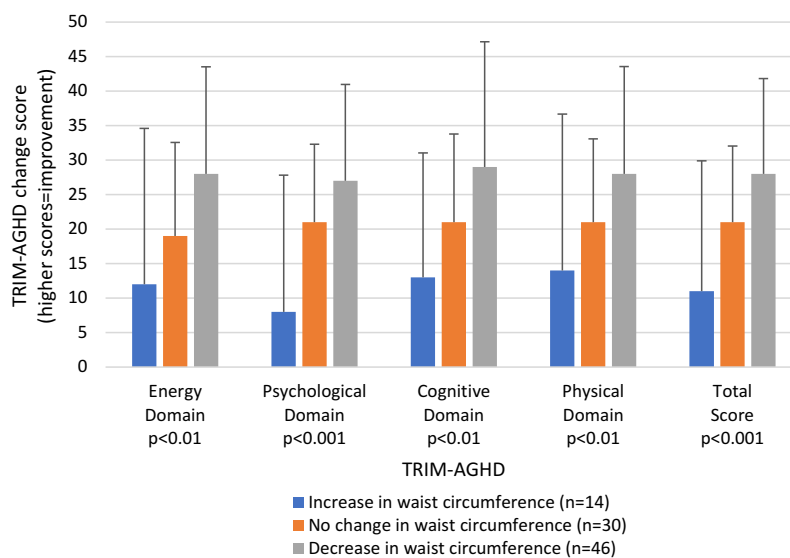
Table 3 TRIM-AGHD change scores by Patient Global Rating of Change (PGRC)

TRIM-AGHD subscale	PGRC			Anchor-based MID ^a
	Better (<i>n</i> = 57) Mean (SD)	About the same (<i>n</i> = 33) Mean (SD)	Worse (<i>n</i> = 6) Mean (SD)	
Energy	– 31.8 (13.8)	– 10.0 (11.5)	7.2 (9.5)	21.80
Psychological	– 30.4 (13.4)	– 10.8 (11.4)	3.6 (11.8)	19.63
Cognitive	– 32.6 (16.6)	– 11.8 (11.4)	– 2.0 (9.0)	20.80
Physical	– 32.0 (15.1)	– 11.8 (10.5)	– 2.5 (8.8)	20.20
Total Score	– 31.4 (12.9)	– 11.0 (8.9)	2.4 (9.0)	20.43

MID minimal important difference, PGRC Patient Global Rating of Change, SD standard deviation, TRIM-AGHD Treatment-Related Impact Measure—Adult Growth Hormone Deficiency

^aMID = difference between ‘Better’ and ‘About the Same’

Fig. 1 TRIM-AGHD change by waist circumference. TRIM-AGHD Treatment-Related Impact Measure—Adult Growth Hormone Deficiency



Note: The change scores of the TRIM-AGHD are absolute changes based on the transformed 0-100 point subscale and total scores.

endpoint to assess this improvement. Incorporating the TRIM-AGHD in routine clinical visits can help both patients and their healthcare providers to better understand the effects of treatment and accurately assess this change over time. However, little work in MID estimation has been done in the context of AGHD and this study represents a step forward in understanding how to interpret the magnitude of change of a given therapy in both the clinical and research setting.

We focused on the evaluation of anchor-based and distribution-based approaches to defining the MID. A limitation in the application of different anchors or anchor types may produce different values of the MID [38], as could be true with the distribution-based methods in which different statistical approaches may also produce differing MIDs. Additional limitations include a potential differing of defined MID values based on whether data collection of the

anchor was prospective versus retrospective [39] or the possibility that the MID as determined by anchor-based methods falls within the instrument’s random variation [23]. Distribution-based methods are limited by their ability to define only a minimal value below which a change in outcome score for a given measure may be due to measurement error [40], which does not provide information on clinical importance.

This study focused on the further evaluation of the TRIM-AGHD with respect to its ability to be sensitive to changes experienced by patients. As seen in this study, the TRIM-AGHD was highly responsive in GHD patients starting GH treatment as evidenced by effect sizes exceeding 1.00 in 93.8% of patients. Distribution-based and anchor-based approaches were used to converge and establish the MID. It should be noted that the estimated MID for the anchor-based approach using patient-reported

perceived change was 20.43 while the distribution-based values were smaller (2.66 for SEM and 8.09 for 0.5 SD), suggesting that even with a small treatment effect, patients are reporting large improvements in functioning and wellbeing. Regarding the relationship between the SEM and the MID, it should be pointed out that while some studies show excellent agreement between SEM and minimal differences, others are not as strong [41]. Our goal here, given the variation between distributional-based and anchor-based approaches, was to triangulate the various methods as no one method is more robust than the other. This way we allowed the different methods to account for an MID estimation until further studies are examined.

When evaluating these results using the COSMIN ‘responsiveness’ checklist [32], the methodology of the study was rated as good or excellent on each of the criteria. However, as with all studies, there are limitations to this one. This study was US-based and data is predominantly from one site, and findings from other countries, especially those with differing cultural beliefs, may be different. While we acknowledge the numbers used in the analysis were small, we believe the sample size was adequate for performing the psychometric test used to evaluate the MID in this study, as the majority (59%) of the sample reported a minimal level of improvement necessary to allow the calculation of an MID for each of the methods used. Additionally, the etiology of the GHD was unknown for almost half the sample, a greater percent than would be expected. This is most likely due to the fact that the information was self-reported by the patient who may not have been aware of their GHD etiology. However, given that the patients were recruited by physicians from their own practice and an inclusion criterion required a confirmed diagnosis of GHD, we do not believe that the unknown etiology is reflective of an unclear diagnosis.

Another limitation revolves around the recent treatments for AGHD and how quickly they act. In this study, treatment benefit was reported by patients soon after starting treatment. In fact, one-third (33.3%) of patients, on their first report of improvement as early as week 4, had more than a minimal amount of improvement. The use of anchor-based approaches in situations like this are not optimal as minimal changes are typically not caught in time, and the more subtle changes are missed. To adjust for this reality, we used the first three levels out of seven possible levels of reported improvement. This may explain why the estimates using the PGRC are larger than the estimates found from the distribution-based approaches.

Additionally, concomitant treatment was not investigated in this study. We believe this would be of value to investigate in future studies. Also, understanding the relationship between patient- and physician-reported MID would be of interest.

5 Conclusions

The suggested MID for the TRIM-AGHD based on this study is an improvement of 10 points in the total score. Improvements that meet or exceed this threshold should be considered clinically relevant and important. Thus, having patients complete the TRIM-AGHD in both a research and/or a clinical setting can be a valuable tool for assessing patient-reported treatment benefit. Given the high degree of responsiveness to treatment of the measure, by applying an MID value of 10 points to interpreting change in total scores, researchers can better assess the full range of differences when comparing treatments and clinicians can better assess if treatment is effective for a given patient.

Author Contributions MB contributed to the research design, analysis, and manuscript preparation; JFB contributed to study design and conduct, and manuscript preparation; DMB conducted the data analysis and contributed to manuscript preparation; JEA, LW, LH, and MHR contributed to research design and manuscript preparation. All authors read and approved the final manuscript. MB is the corresponding author and overall guarantor.

Data Availability Statement The datasets generated and/or analyzed during the current study are not publicly available due to being proprietary but are available from the corresponding author on reasonable request and dependent upon purpose.

Compliance with Ethical Standards

Funding This study was funded by Novo Nordisk A/S.

Conflict of interest MB, JFB, and DMB are advisors/paid consultants to Novo Nordisk. LH, MHR, and LW are employees of Novo Nordisk. JEA was an employee of Novo Nordisk at the time of the study.

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 and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Ethics approval was granted by Copernicus Group IRB (Tracking # TBG1-13-475).

Informed consent Informed consent was obtained from all individual participants included in the study.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Monson JP, Brooke AM, Akker S. Adult Growth Hormone Deficiency. [Updated 2015 May 19]. In: De Groot LJ, Chrousos G, Dungan K, et al., eds. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000. <https://www.ncbi.nlm.nih.gov/books/NBK278982/> Accessed 01 Sep 2017.
2. Human Growth Foundation. Adult Growth Hormone Deficiency. http://hgfound.org/resources/adult-growth-hormone-deficiency/#res_aghd.explained3. Accessed 7 June 2016.
3. Kaufman JM, Taelman P, Vermeulen A, Vandeweghe M. Bone mineral status in growth hormone-deficient males with isolated and multiple pituitary deficiencies of childhood onset. *J Clin Endocrinol Metab.* 1992;74:118–23.
4. Laursen T, Jorgensen JO, Christiansen JS. The management of adult growth hormone deficiency syndrome. *Expert Opin Pharmacother.* 2008;9:2435–50.
5. Gazzaruso C, Gola M, Karamouzis I, Giubbini R, Giustina A. Cardiovascular risk in adult patients with growth hormone (GH) deficiency and following substitution with GH—an update. *J Clin Endocrinol Metab.* 2014;99:18–29.
6. Christensen TL, Djurhuus CB, Clayton P, Christiansen JS. An evaluation of the relationship between adult height and health-related quality of life in the general UK population. *Clin Endocrinol (Oxf).* 2007;67:407–12.
7. Coste J, Pouchot J, Carel JC. Height and health-related quality of life: a nationwide population study. *J Clin Endocrinol Metab.* 2012;97:3231–9.
8. Chrisoulidou A, Kousta E, Beshyah SA, Robinson S, Johnston DG. How much, and by what mechanisms, does growth hormone replacement improve the quality of life in GH-deficient adults? *Baillieres Clin Endocrinol Metab.* 1998;12:261–79.
9. McGauley GA. Quality of life assessment before and after growth hormone treatment in adults with growth hormone deficiency. *Acta paediatrica Scandinavica Supplement.* 1989;356:70–2 (**discussion 3–4**).
10. Rosen T, Wiren L, Wilhelmsen L, Wiklund I, Bengtsson BA. Decreased psychological well-being in adult patients with growth hormone deficiency. *Clin Endocrinol (Oxf).* 1994;40:111–6.
11. Mahajan T, Crown A, Checkley S, Farmer A, Lightman S. Atypical depression in growth hormone deficient adults, and the beneficial effects of growth hormone treatment on depression and quality of life. *Eur J Endocrinol.* 2004;151:325–32.
12. Stouthart PJ, Deijen JB, Roffel M, Delemarre-van de Waal HA. Quality of life of growth hormone (GH) deficient young adults during discontinuation and restart of GH therapy. *Psychoneuroendocrinology.* 2003;28:612–26.
13. McMillan CV, Bradley C, Gibney J, Healy ML, Russell-Jones DL, Sonksen PH. Psychological effects of withdrawal of growth hormone therapy from adults with growth hormone deficiency. *Clin Endocrinol (Oxf).* 2003;59:467–75.
14. Malik IA, Foy P, Wallymahmed M, Wilding JP, MacFarlane IA. Assessment of quality of life in adults receiving long-term growth hormone replacement compared to control subjects. *Clin Endocrinol (Oxf).* 2003;59:75–81.
15. Burman P, Broman JE, Hetta J, Wiklund I, Erfurth EM, Hagg E, et al. Quality of life in adults with growth hormone (GH) deficiency: response to treatment with recombinant human GH in a placebo-controlled 21-month trial. *J Clin Endocrinol Metab.* 1995;80:3585–90.
16. Cuneo RC, Judd S, Wallace JD, Perry-Keene D, Burger H, Lim-Tio S, et al. The Australian Multicenter Trial of Growth Hormone (GH) Treatment in GH-Deficient Adults. *J Clin Endocrinol Metab.* 1998;83:107–16.
17. Ikeda H, Kudo M. Long-term follow-up results of growth hormone therapy for patients with adult growth hormone deficiency. *Hormones (Athens, Greece).* 2016;15:45–53.
18. U.S. Dept. of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research: Center for Biologics Evaluation and Research: Center for Devices and Radiological Health. (2009). Guidance for industry: patient-reported outcome measures, use in medical product development to support labeling claims. <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>. Accessed 15 Nov 2016.
19. Brod M, Hojbjerg L, Adalsteinsson JE, Rasmussen MH. Assessing the impact of growth hormone deficiency and treatment in adults: development of a new disease-specific measure. *J Clin Endocrinol Metab.* 2014;99:1204–12.
20. Yost KJ, Eton DT. Combining distribution- and anchor-based approaches to determine minimally important differences: the FACIT experience. *Eval Health Prof.* 2005;28:172–91.
21. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes.* 2006;4:70.
22. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61:102–9.
23. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol.* 2003;56:395–407.
24. 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;77:371–83.
25. Wyrwich KW, Bullinger M, Aaronson N, Hays RD, Patrick DL, Symonds T, et al. Estimating clinically significant differences in quality of life outcomes. *Qual Life Res.* 2005;14:285–95.
26. Shikar R, Harding G, Leahy M, Lennox RD. Minimal important difference (MID) of the Dermatology Life Quality Index (DLQI): results from patients with chronic idiopathic urticaria. *Health Qual Life Outcomes.* 2005;3:36.
27. Leidy NK, Wyrwich KW. Bridging the gap: using triangulation methodology to estimate minimal clinically important differences (MCIDs). *COPD.* 2005;2:157–65.
28. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertain the minimal clinically important difference. *Control Clin Trials.* 1989;10:407–15.
29. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol Drug Saf.* 2008;17:200–8.
30. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310:2191–4.
31. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012;21:651–7.
32. Consensus-based Standards for the Selection of Health Measurement Instruments. COSMIN Checklist. 2011. <http://www.cosmin.nl/images/upload/files/COSMIN%20checklist%20with%204-point%20scale%2022%20juni%202011.pdf>. Accessed 15 Mar 2018.
33. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health-status. *Med Care.* 1989;27:S178–89.
34. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale: L. Erlbaum Associates; 1988.

35. Farivar SS, Liu H, Hays RD. Half standard deviation estimate of the minimally important difference in HRQOL scores? *Expert Rev Pharmacoecon Outcomes Res.* 2004;4:515–23.
36. Rejas J, Pardo A, Ruiz MA. Standard error of measurement as a valid alternative to minimally important difference for evaluating the magnitude of changes in patient-reported outcomes measures. *J Clin Epidemiol.* 2008;61:350–6.
37. Traub RE, Rowley GL. Understanding reliability. *Educ Meas Issues Pract.* 1991;10:37–45.
38. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J.* 2007;7:541–6.
39. Hudgens SA, Yost K, Cella D, Hahn EA, Peterman A. Comparing retrospective and prospective anchors for identifying minimally important differences [abstract]. *Qual Life Res.* 2002;11:629.
40. Gatchel RJ, Lurie JD, Mayer TG. Minimal clinically important difference. *Spine (Phila, Pa 1976).* 2010;35:1739–43.
41. Wyrwich KW. Minimal important difference thresholds and the standard error of measurement: is there a connection? *J Biopharm Stat.* 2004;14:97–110.