Br J Dermatol 2025; **00**:1–12
https://doi.org/10.1093/bjd/ljaf067
Advance access publication date: 2 April 2025

BJD
British Journal of Dermatology
Outcomes and Qualitative Research

Psychometric validation and interpretation thresholds of the Hidradenitis Suppurativa Quality of Life (HiSQOL©) questionnaire using pooled data from the phase III BE HEARD I & II trials of bimekizumab in hidradenitis suppurativa

Joslyn S Kirby,[1] Linnea Thorlacius,[2] Jérémy Lambert,[3] Valerie Ciaravino,[3] Robert Rolleri,[4] Ingrid Pansar,[5] Edward Muller,[6] Christopher G Pelligra[7] and John R Ingram[8]

[1]Department of Dermatology, Penn State University, Hershey, PA, USA
[2]Department of Dermatology, Zealand University Hospital, Roskilde, Denmark; Health Sciences Faculty, University of Copenhagen, Copenhagen, Denmark
[3]UCB Pharma, Colombes, France
[4]UCB Pharma, Morrisville, NC, USA
[5]UCB Pharma, Brussels, Belgium
[6]UCB Pharma, Slough, UK
[7]Evidera, Atlanta, GA, USA
[8]Department of Dermatology & Academic Wound Healing, Division of Infection and Immunity, Cardiff University, Cardiff, UK
Correspondence: Joslyn S. Kirby. Email: jsciacca03@hotmail.com

Abstract

Background Hidradenitis suppurativa (HS) is a chronic, inflammatory skin condition which negatively impacts patients' physical and mental wellbeing. The HS Quality of Life questionnaire (HiSQOL©) was developed to assess HS-specific changes in health-related quality of life (HRQoL), one of the six domains of the core outcome set, established by the HS Core Outcomes Set International Collaboration.

Objectives To evaluate the psychometric properties of HiSQOL total and domain scores and determine interpretation thresholds to guide score interpretation.

Methods Blinded, pooled data from two bimekizumab phase III trials (BE HEARD I & II) in adult patients with moderate-to-severe HS were used to conduct a confirmatory factor analysis (CFA) and to assess convergent validity [correlations with other patient-reported outcomes (PROs)], known-groups validity [grouping patients according to Hurley stage, International Hidradenitis Suppurativa Severity Score System (IHS4) and patient global impression of HS severity scale], reliability [Cronbach's α and intraclass correlation coefficients (ICCs)] and responsiveness (sensitivity to change). Clinically meaningful within-patient improvement thresholds were estimated by anchor- and distribution-based analyses. Symptom/impact severity thresholds were estimated by receiver operating characteristic curve analyses.

Results The CFA models supported the relevance of the three subscales and an underlying unidimensional concept, validating the total score derivation from all items. HiSQOL subscale and total scores showed good convergent and known-groups validity, with HiSQOL scores being consistently higher (worse HRQoL outcomes) for patients with higher disease severity. HiSQOL demonstrated good internal consistency (Cronbach's $\alpha \geq 0.81$ for all scores) and test–retest reliability (ICC 0.73–0.83 across HiSQOL scores). Correlation coefficients between changes in HiSQOL scores and changes in other PRO scores were all positive and statistically significant (P-values < 0.001), with most exceeding 0.30, demonstrating acceptable responsiveness. Clinically meaningful within-patient improvement thresholds were estimated as: 20–21-point decrease for HiSQOL total score (total possible range 0–68), 5–6-point decrease for symptoms (range 0–16), 4–5-point decrease for psychosocial (range 0–20) and 10–11-point decrease for activities–adaptations (range 0–32). Thresholds for different levels of symptom/impact severity were derived.

Conclusions The HiSQOL subscale and total scores demonstrated robust psychometric properties, supporting the use of HiSQOL to interpret trial results and inform treatment decisions.

Accepted: 17 February 2025

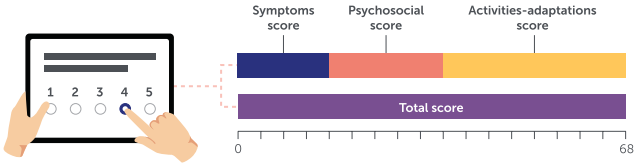
Graphical Abstract

Psychometric Validation and Interpretation Thresholds of the Hidradenitis Suppurativa Quality of Life (HiSQOL©) Questionnaire Using Pooled Data from the Phase III BE HEARD I & II Trials of Bimekizumab in Hidradenitis Suppurativa

Introduction

Hidradenitis suppurativa (HS) is a chronic, debilitating, inflammatory skin disease characterized by painful skin lesions.

The **Hidradenitis Suppurativa Quality of Life (HiSQOL©) questionnaire** is a 17-item patient-reported outcome measure developed to assess HS-specific impacts on patients' lives. A total score and three subscale scores (symptoms, psychosocial, and activities-adaptations) can be derived.



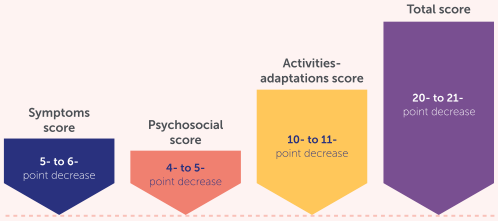
Higher scores correspond to **higher symptomology or impact** on health-related quality of life.

Methods

Psychometric performance of the HiSQOL total and subscale scores were assessed and thresholds for **clinically meaningful within-patient improvements** and for different symptom/impact severity levels were determined using pooled, blinded data from **two phase III trials (BE HEARD I & II)** evaluating bimekizumab in patients with **moderate to severe HS.**

Results

- HiSQOL demonstrated:
  - ☑ Good convergent and known-groups validity
  - ☑ Good sensitivity to change
  - ☑ Good test-retest and internal consistency reliability
- Thresholds of none, mild, moderate, severe and very severe symptom/impact **severity levels** were identified.
- Thresholds for **clinically meaningful within patient improvement** are the following:



Conclusions

The HiSQOL subscale and total scores are sufficiently **reliable and valid**. Along with the derived interpretation thresholds, these scores can therefore be trusted to **monitor HS-specific health-related quality of life** and inform treatment outcome decisions.

Lay summary

Hidradenitis suppurativa (HS) is a long-term skin disease that causes painful lesions. HS affects people physically and psychologically. This can make it difficult to do activities of daily living, such as walking or getting dressed. Currently, there are few medications that people can take to treat HS. We want to understand how medications may help people with HS. To do this, we need to consider patients' physical, emotional and social functioning. These outcomes can be measured using self-completed questionnaires. There has been a lack of HS-specific quality-of-life questionnaires. Thus, the patient-reported Hidradenitis Suppurativa Quality of Life questionnaire (HiSQOL©) was created. The HiSQOL asks patients with HS 17 questions to understand how they feel over a 1-week period. There were two phase III clinical trials studying patients with moderate-to-severe HS. Patients in these studies completed the HiSQOL.

We carried out a series of statistical analyses in this study. This confirmed the validity, reliability and robustness of the data obtained from the HiSQOL. This questionnaire assessed patients' physical, emotional and social functioning. It was sensitive enough to reveal changes over time. It could distinguish between patient groups with different levels of HS severity. Clinicians need to determine whether an improvement in a patient's HiSQOL score is meaningful to the patient. We established thresholds to help with this.

Overall, our findings suggest the HiSQOL could be a useful tool for informing treatment choices for people living with HS.

What is already known about this topic?

- Hidradenitis suppurativa (HS) is a chronic, inflammatory disease that causes painful skin lesions, drainage and fatigue, impacting patients' quality of life (QoL).
- While general health-related QoL questionnaires and pan-dermatology QoL instruments exist, these are limited in assessing specific aspects of HS, such as pain and drainage.

What does this study add?

- The Hidradenitis Suppurativa Quality of Life questionnaire (HiSQOL©) is valid, reliable and sensitive to change over time, making it a fit-for-purpose HS-specific QoL instrument.

What are the clinical implications of the work?

- By focusing on HS-specific symptoms, activities–adaptations and psychosocial consequences that impact daily activities and well-being, the HiSQOL©, with its interpretation thresholds, provides a clinically meaningful tool to capture a holistic view of the disease beyond the lesions for patients, treating physicians and patient organizations.

Hidradenitis suppurativa (HS) is a chronic, relapsing and debilitating inflammatory skin disease characterized by painful skin lesions.[1–3] Beyond high pain levels, patients experience difficult symptoms, including drainage, odour and fatigue.[1–3] HS symptoms can impact daily activities and psychosocial aspects of patients' lives.[2–5]

Changes in patients' HS symptoms and physical, emotional and social functioning are key factors in the holistic evaluation of treatment efficacy for moderate-to-severe HS.[6] A core outcome set of domains (i.e. an agreed minimum set of measurements in all clinical trials for HS) has been established by the Hidradenitis SuppuraTiva cORe outcomes set International Collaboration (HiSTORIC) using a Delphi process involving patients and healthcare providers.[7] This set includes the concurrent measurement of five domains, including HS-specific health-related quality of life (HRQoL).[7]

While dermatology-specific HRQoL instruments exist [e.g. Dermatology Life Quality Index (DLQI) and Skindex], the items covered are broad-ranging and may not capture HS-specific features.[8–10] Thus, the Hidradenitis Suppurativa Quality of Life questionnaire (HiSQOL©) was developed collaboratively with patient research partners, researchers with expertise in instrument development and clinicians with expertise in HS, to assess changes in HS-specific HRQoL in clinical trials.[9,11]

The psychometric properties of a new measure should be assessed in the relevant patient population to ensure it is fit-for-purpose.[12–14] In addition, the derivation of severity thresholds and clinically meaningful within-patient thresholds helps guide interpretation of absolute scores and changes in scores, respectively.[12–14]

Here, we assessed the psychometric properties of and derived interpretation thresholds for HiSQOL subscale and total scores, using data from two phase III trials (BE HEARD I & II) evaluating bimekizumab in moderate-to-severe HS.

Materials and methods

BE HEARD I & II

Psychometric analyses were conducted using blinded, pooled data from two phase III trials, BE HEARD I (NCT04242446) and II (NCT04242498).[15,16] Patients with moderate-to-severe HS received (initial weeks 0–16/maintenance weeks 16–48) bimekizumab 320 mg every 2 weeks (Q2W)/Q2W, bimekizumab Q2W/every 4 weeks (Q4W), bimekizumab Q4W/Q4W or placebo/bimekizumab Q2W. Further details are included in Methods S1 (see Supporting Information) and full inclusion/exclusion criteria have been previously published.[17]

HiSQOL

The HiSQOL is a 17-item, HS-specific questionnaire specifically designed to assess HRQoL in patients with HS with a 7-day recall period. Item scores range from 0 to 4 (Table S1; see Supporting Information).

Three subscales are derived as the sum of items pertaining to the subscale: symptoms subscale (four items, score range 0–16), psychosocial subscale (five items, score range 0–20) and activities–adaptations subscale (eight items, score range 0–32). Item scores are summed to create a total score (range 0–68). A higher score indicates a more severe impact on HRQoL.

Other assessments

Four single-item patient-global impression (PGI) scales, assessed at weeks 4 and 16, captured the PGI of HS severity (PGI-S-HS, additionally assessed at baseline), PGI of change in HS severity since starting study medication (PGI-C-HS), PGI of severity of skin pain (PGI-S-SP, additionally assessed at baseline) and PGI of change in severity of skin pain since starting study medication (PGI-C-SP). Further details are included in Methods S1. The DLQI, a 10-item skin disease-specific patient-reported questionnaire to evaluate patients' perceptions of how symptoms and treatment affect their HRQoL, was also assessed.

Clinician-rated assessments based on HS lesions included the International Hidradenitis Suppurativa Severity Score System (IHS4) and the HS Physician's Global Assessment (PGA). Hurley stage categorization was used to stage the severity of the worst affected skin region at these same timepoints.[18]

Psychometric analyses

Psychometric analyses were conducted on the blinded HiSQOL analysis set (all randomized patients who had ≥ 1 non-missing HiSQOL subscale score at any scheduled assessment visit).

A confirmatory factor analysis (CFA) was conducted for HiSQOL at baseline to confirm its structure. A bi-factor model (i.e. each item aligns with one of the three HiSQOL subscales and HiSQOL total score) for the HiSQOL was evaluated. A three-factor structure model (i.e. each item aligns with only one HiSQOL subscale) for the HiSQOL and a one-factor structure model (i.e. each item aligns only with HiSQOL total score) for the HiSQOL were also explored. For all CFA models, the following fit statistics were computed: comparative fit index (CFI), root mean square error of approximation (RMSEA) along with 90% confidence interval (CI) and standardized root mean squared residual (SRMR). The accepted criteria used to aid interpretation are presented in Table S2 (see Supporting Information).[19–23]

Convergent and known-groups validity analyses were conducted to assess the construct validity of the HiSQOL and establish how well it captured the underlying concept

it intended to measure. Regarding convergent validity, Pearson's and Spearman's rank correlation coefficients and corresponding *P*-values were calculated to assess the strength of associations between HiSQOL subscale and total scores vs. measures assessing similar (patient-reported DLQI) and different (IHS4) concepts at baseline and week 16. Correlation strength between two variables was interpreted as weak/moderate/strong/very strong when the correlation coefficient was $< 0.3/\geq 0.3$ to $< 0.7/\geq 0.7$ to $< 0.9/\geq 0.9$, respectively.[24] The following hypotheses for the expected correlations were prespecified in the statistical analysis plan: HiSQOL symptom subscale scores and the total score of the HiSQOL to be moderately or strongly correlated with the IHS4 score, with the psychosocial and activities–adaptations subscale scores of the HiSQOL having relatively smaller correlations with the IHS4; and all HiSQOL subscale scores and the total score of the HiSQOL to be moderately or strongly correlated with the DLQI total score.

Known-groups validity was assessed by analysis of variance (ANOVA) comparing mean HiSQOL scores between groups of patients having different clinical status defined using Hurley stage, IHS4, PGI-S-HS and HS PGA at baseline and week 16.

Internal consistency reliability, showing the extent to which individual items within an instrument correlate with each other to form a multi-item scale, was evaluated using standardized Cronbach's α at baseline and week 16. Standardized α coefficients ≥ 0.70 were regarded as indicative of acceptable/good internal consistency.[25]

Test–retest reliability (i.e. repeatability) was evaluated using intraclass correlation coefficients (ICCs, using a two-way mixed-effect ANOVA model with week as a fixed effect), between baseline and week 4 and between weeks 32 and 48, in stable patients (defined as those with no change in PGI-S-HS score over the respective time intervals). An ICC value > 0.70 was considered acceptable.[19]

Responsiveness (i.e. sensitivity to change over time) was evaluated by correlating changes from baseline to week 16 in HiSQOL subscale and total scores with changes in PGI scales within that same time interval, using Spearman's rank correlation coefficient. A threshold of 0.30 Spearman's rank correlation was considered to demonstrate acceptable sensitivity to change over time.[26]

Interpretation thresholds

Anchor- and distribution-based analyses were used to determine clinically meaningful within-patient improvement thresholds. In the anchor-based analysis, patients were classified into response groups according to PGI-S-HS and PGI-C-HS. PGI-C-HS was considered as a supportive anchor only as it requires a patient to recall their status from baseline, and thus is potentially subject to recall bias. Descriptive statistics were calculated for HiSQOL score change from baseline to week 16 within these groups. A two-category improvement for PGI-C-HS defined a patient as 'much better'. Empirical cumulative distribution function (eCDF) and probability density function (using Kernel density estimation) curves of changes from baseline at week 16, in each HiSQOL subscale/total score, were plotted separately for each response group within each selected anchor. In the distribution-based analysis, SEM and 0.5 SD were calculated using baseline data. Estimates from anchor- and distribution-based analyses were triangulated to determine a range of thresholds for clinically meaningful within-patient change. Greater weight was given to anchor-based vs. distribution-based estimates.

Additionally, thresholds for different severity levels were derived for each HiSQOL subscale/total score using PGI-S-HS severity levels (none, mild, moderate, severe, very severe) as anchors. To determine the severity thresholds for a given target scale, a receiver operating characteristic (ROC) curve analysis was employed. The optimal threshold for a given severity level was estimated as that which maximized the Youden Index, using data pooled across visits at baseline and weeks 4, 16, 32 and 48. The value for the area under the curve (AUC) of a ROC curve of 0.50 indicates that the use of a given target scale is no better at differentiating responder and nonresponder groups than random chance, whereas a value of ≥ 0.70 indicates a satisfactory level of accuracy in differentiating those groups.[27]

Results

Patient disposition and baseline characteristics

Baseline characteristics were collected from 1010 patients enrolled in BE HEARD I ($N=502$) and II ($N=508$). The mean (SD) age at baseline was 36.7 (12.2) years, with the largest proportion of patients being women (56.6%) and White (76.3%). At baseline, 59.2% of patients had a body mass index of ≥ 30 kg m^{-2}. Overall, patients had a mean (SD) HS disease duration of 8.0 (7.8) years. The proportions of patients with Hurley stage II and III were 55.6% and 44.4%, respectively (Table 1).[28] At baseline, 997 of 1010 patients had nonmissing HiSQOL scores (1.3% missing). The mean (SD) HiSQOL total score was 25.2 (13.4). No significant floor or ceiling effects were noted at baseline (Table S3; see Supporting Information).

Confirmatory factor analysis

The bi-factor model showed better model fit than three-factor and one-factor structure models. CFI values were 0.91 (meeting the criterion for acceptable model fit), 0.82 and 0.76 for the bi-factor, three-factor and one-factor (unidimensional) models, respectively (Table S2). SRMR residual values were 0.05 for the bi-factor model, 0.07 for the three-factor model and 0.08 for the one-factor (unidimensional) model, all meeting the criterion for acceptable model fit. RMSEA values were higher than the acceptable criterion of 0.08 for all models [although the RMSEA value (0.084) for the bi-factor model was close to meeting the acceptability criterion]. In the bi-factor model, all items had loadings > 0.40 with the general factor (representing the total score) while they had weak loadings with their respective group factor (representing their respective subscale). In the three-factor structure model, all items had adequate loadings (> 0.40) onto their hypothesized group factor. In the one-factor (unidimensional) structure model, all items had adequate loadings (> 0.40) onto the single factor, providing

Table 1 Baseline demographic and disease characteristics of patients

Characteristic	HiSQOL analysis set (*N* = 1010)
Age (years), mean (SD)	36.7 (12.2)
Women, *n* (%)	572 (56.6)
Race, *n* (%)	
White	771 (76.3)
Black or African American	106 (10.5)
Asian	41 (4.1)
Other or Mixed[a]	46 (4.6)
Missing	46 (4.6)
Region, *n* (%)	
North America	385 (38.1)
Western Europe	290 (28.7)
Central and Eastern Europe	260 (25.7)
Asia and Australia	75 (7.4)
Body mass index (kg m^{-2}), *n* (%)	
< 25	156 (15.4)
25 to < 30	253 (25.0)
≥ 30[b]	598 (59.2)
Missing	3 (0.3)
Duration of disease (years), mean (SD)	8.0 (7.8)
Hurley stage, *n* (%)[c]	
I	0
II	562 (55.6)
III	448 (44.4)
HiSQOL baseline scores, mean (SD)	
Symptoms score (range 0–16)	7.9 (3.5)
Psychosocial score (range 0–20)	5.3 (4.4)
Activities–adaptations score (range 0–32)	12.0 (7.2)
Total score (range 0–68)	25.2 (13.4)

[a]Includes American Indian or Alaska Native (*n* = 2, 0.2%), Native Hawaiian or other Pacific Islander (*n* = 2, 0.2%). [b]Classed as clinically obese.[28] [c]Only patients with Hurley stage II and III were included at baseline, as per the BE HEARD I & II inclusion and exclusion criteria. HiSQOL, Hidradenitis Suppurativa Quality of Life questionnaire.

support for a single factor based on total score, even though its model fit was suboptimal (Table S2).

Construct validity

At baseline, all correlations with the IHS4 and DLQI total score were positive and significant. A weak correlation (r < 0.30) was observed between IHS4 scores and all subscale scores and total score. A moderate correlation was observed between HiSQOL symptoms subscale score and DLQI total score (r = 0.66), while a strong correlation was observed between DLQI and HiSQOL total scores, psychosocial and activities–adaptations subscale scores (r > 0.70, Table S4; see Supporting Information). At week 16, the correlation between psychosocial subscale score and IHS4 was weak (r = 0.28), while symptoms subscale score, activities–adaptations subscale score and total score were moderate (r = 0.46, r = 0.37 and r = 0.42, respectively). A strong correlation was observed between DLQI total score and all HiSQOL subscale scores (r > 0.70, Table S4).

HiSQOL subscale and total scores showed known-groups validity based on Hurley stage and IHS4 (Figure 1). Mean HiSQOL scores were consistently higher (worse HRQoL outcomes) for patients with higher disease severity. Of note, at week 16, for the psychosocial subscale score, IHS4 was similar between mild and moderate, but differed between moderate and severe. All distributions of HiSQOL subscale and total scores significantly differed by known-groups, as shown by the *P*-values from the Kruskal–Wallis test ($P < 0.05$, Figure 1). Similar results were observed using other patient-reported (PGI-HS) and clinician-rated (HS PGA) anchors (data not shown).

Reliability

Good internal consistency reliability was established for the HiSQOL. When assessed with standardized Cronbach's α, at baseline, all subscale and total scores indicated a good internal consistency; this was also observed at week 16 (Table 2).[19,25]

Test–retest reliability was also acceptable for all HiSQOL subscale scores and the total score. ICC values were 0.73–0.78 between baseline and week 4 and 0.76–0.83 between weeks 32 and 48 (Table 2).[19,25]

Responsiveness

All correlation coefficients were positive (as expected) and statistically significant (*P*-values < 0.001). Most correlations, except for PGI-C-HS and psychosocial subscale score and PGI-C-SP and psychosocial subscale from baseline to week 16 (0.24 and 0.25, respectively), exceeded the threshold of 0.30 to demonstrate acceptable sensitivity. Correlation coefficients between changes in the PGI-S-HS from baseline to week 16 and changes in the symptoms subscale, activities–adaptations subscale and total scores all exceeded 0.50, indicating moderate-to-high responsiveness (Table 3).

Interpretation thresholds

Clinically meaningful within-patient change thresholds

Patients with a one- and two-level improvement on the PGI-S-HS from baseline to week 16 are shown in Table 4. The findings from the eCDF plots supported the use of estimates from the group with two levels of improvement on the PGI-S-HS as an anchor to derive clinically meaningful within-patient change thresholds, due to the larger degree of separation between no change and two levels of improvement, compared with no change and one level (Figure 2). PGI-C-HS, considered as a supportive anchor only as it requires a patient to recall their status from baseline, provided further evidence on the selection of the thresholds (Table S5, Figure S1; see Supporting Information).

Triangulation of various estimates from anchor-based and distribution-based approaches indicated that clinically meaningful within-patient improvement thresholds should be a 20–21-point decrease for HiSQOL total score. For HiSQOL subscale scores, the following clinically meaningful within-patient improvement thresholds were identified: symptoms, 5–6-point decrease; psychosocial, 4–5-point decrease; activities–adaptations, 10–11-point decrease. Due to the small sample size of patients whose HS was worsening and a very small effect size, clinically meaningful within-patient change thresholds for worsening were not derived or recommended.

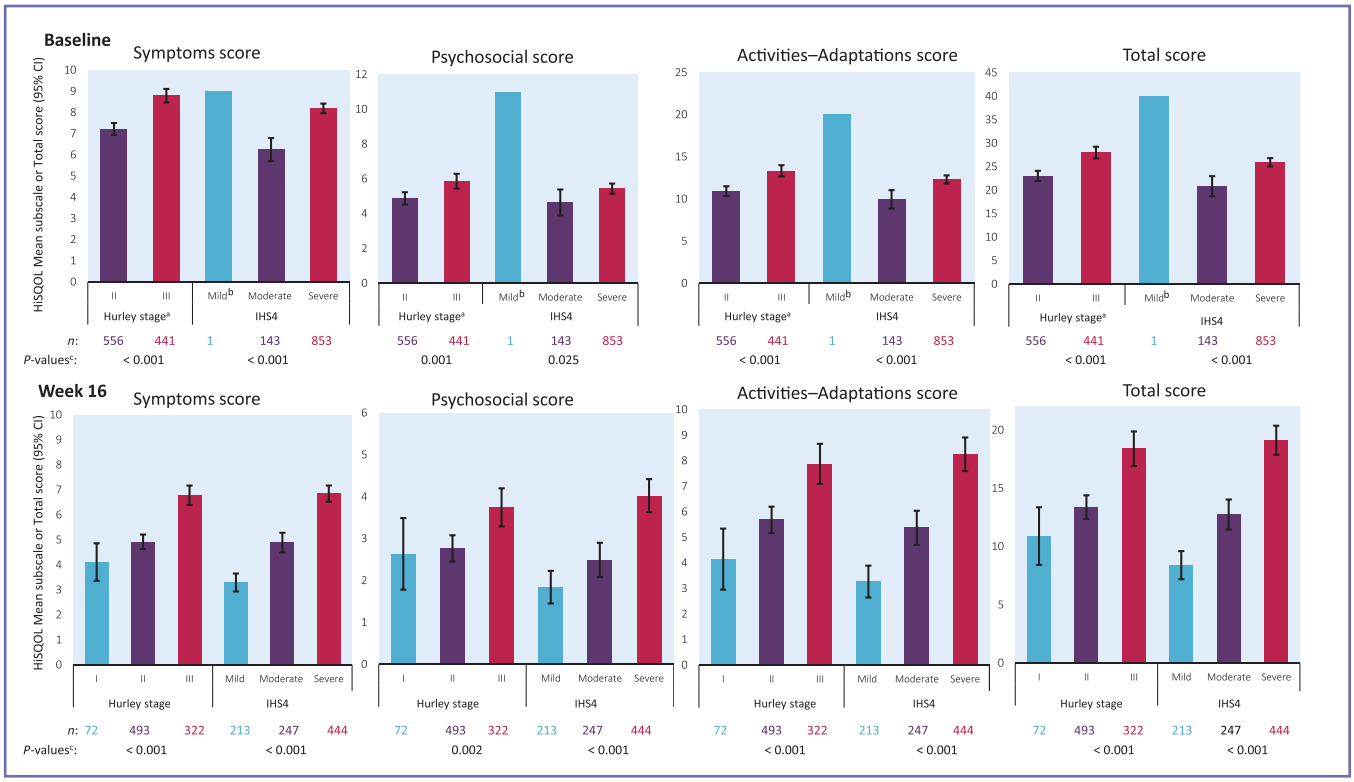


Figure 1 HiSQOL subscale/total scores in HS severity groups as defined by Hurley stage and IHS4 severity at baseline (upper) and week 16 (lower) (known-groups validity).[a]Only patients with Hurley stage II and III were included at baseline, as per the BE HEARD I & II inclusion and exclusion criteria. [b]Only one patient had an IHS4 score of mild at baseline. [c]Overall *P*-values estimated from the Kruskal–Wallis test. CI, confidence interval; HiSQOL, Hidradenitis Suppurativa Quality of Life questionnaire; HS, hidradenitis suppurativa; IHS4, International HS Severity Score System.

Table 2 Internal consistency and test–retest reliability for HiSQOL subscale and total scores

	Internal consistency reliability, standardized Cronbach's α[a]		Test–retest reliability, ICC[b]	
HiSQOL subscale/total score	**Baseline (*n*=997)**	**Week 16 (*n*=908)**	**Baseline vs. Week 4 (*n*=356)**	**Week 32 vs. Week 48 (*n*=265)**
HiSQOL symptoms score	0.81	0.84	0.73	0.77
HiSQOL psychosocial score	0.81	0.84	0.75	0.76
HiSQOL activities–adaptations score	0.86	0.90	0.73	0.79
HiSQOL total score	0.92	0.94	0.78	0.83

[a]Standardized α coefficients of ≥0.70 are indicative of acceptable or good internal consistency.[25] [b]ICC > 0.70 indicates acceptable test–retest reliability.[19] HiSQOL, Hidradenitis Suppurativa Quality of Life questionnaire; ICC, intraclass correlation coefficients.

Table 3 Correlation between changes in HiSQOL subscale and total scores and changes in external anchors at week 16 (responsiveness)

	External anchors			
Spearman's correlation,[a,b] r	**PGI-S-SP**	**PGI-C-SP**	**PGI-S-HS**	**PGI-C-HS**
n	896	898	895	898
HiSQOL symptoms score	0.63	0.43	0.63	0.43
HiSQOL psychosocial score	0.36	0.25[c]	0.41	0.24[c]
HiSQOL activities–adaptations score	0.52	0.35	0.54	0.35
HiSQOL total score	0.58	0.39	0.61	0.39

[a]$P<0.001$ for all. [b]All values show a moderate correlation ($0.30 \leq r < 0.70$), except for those marked. [c]Weak correlation ($r < 0.30$). HiSQOL, Hidradenitis Suppurativa Quality of Life questionnaire; PGI-C-HS, patient global impression of change in hidradenitis suppurativa severity; PGI-C-SP, patient global impression of change in severity of skin pain; PGI-S-HS, patient-global impression of HS severity; PGI-S-SP, patient global impression of severity of skin pain.

Table 4 Anchor-based clinically meaningful within-patient change threshold: observed changes from baseline at week 16 by PGI-S-HS[a]

	Improvement by				
Statistic	**4 levels**	**3 levels**	**2 levels**	**1 level**	**No change**
CFB	−4	−3	−2	−1	0
HiSQOL symptoms score[b]					
n	–	32	143	336	288
Mean (95% CI[c])	–	−7.47 (−8.64 to −6.30)	−5.56 (−6.08 to −5.04)	−2.98 (−3.25 to −2.71)	−0.84 (−1.14 to −0.54)
Median (IQR)	–	−7.00 (−10.50 to −5.00)	−6.00 (−8.00 to −4.00)	−3.00 (−4.00 to −1.00)	−1.00 (−3.00 to 1.00)
HiSQOL psychosocial score[b]					
n	–	32	143	336	288
Mean (95% CI)	–	−4.78 (−6.36 to −3.21)	−4.38 (−5.11 to −3.64)	−2.66 (−3.02 to −2.31)	−1.19 (−1.55 to −0.84)
Median (IQR)	–	−4.00 (−8.50 to −1.00)	−4.00 (−7.00 to −1.00)	−2.00 (−4.50 to 0.00)	−1.00 (−2.00 to 0.00)
HiSQOL activities–adaptations score[d]					
n	2	32	143	336	288
Mean (95% CI)	−26.00 (−64.12 to 12.12)	−12.09 (−14.48 to −9.71)	−10.87 (−11.94 to −9.79)	−6.60 (−7.20 to −6.00)	−2.97 (−3.55 to −2.39)
Median (IQR)	−26.00 (−29.00 to −23.00)	−10.50 (−16.00 to −7.00)	−11.00 (−16.00 to −7.00)	−6.00 (−10.00 to −3.00)	−3.00 (−6.00 to 0.00)
HiSQOL total score[d]					
n	2	32	143	336	288
Mean (95% CI)	−48.00 (−162.36 to 66.36)	−24.34 (−28.51 to −20.18)	−20.80 (−22.75 to −18.86)	−12.24 (−13.25 to −11.23)	−5.00 (−5.98 to −4.02)
Median (IQR)	−48.00 (−57.00 to −39.00)	−22.50 (−32.00 to −17.00)	−21.00 (−29.00 to −14.00)	−11.00 (−18.00 to −6.00)	−5.00 (−10.00 to 0.00)

[a]The HiSQOL analysis set is defined as all randomized study participants from both studies who had ≥ 1 nonmissing subscale score of the HiSQOL at any scheduled assessment visit. [b]Categories with no more than two study participants were excluded from the table. [c]Due to the small sample size of some groups, some CIs are wide and may fall outside the range of plausible values. [d]Categories with no study participants were excluded from the table. CFB, change from baseline; CI, confidence interval; HiSQOL, Hidradenitis Suppurativa Quality of Life questionnaire; IQR, interquartile range of Q1–Q3; PGI-S-HS, patient-global impression of HS severity.

Severity thresholds

Severity thresholds for total score and all subscales are shown in Table 5. AUC statistics were > 0.70. Detailed statistics, including the Youden Index and AUC, are shown in Table S6 (see Supporting Information).

Discussion

This psychometric analysis demonstrated robust properties of HiSQOL total and subscale scores, including convergent and known-groups validity, sensitivity to change and reliability, in a substantial representative sample of patients with moderate-to-severe HS in BE HEARD I & II.

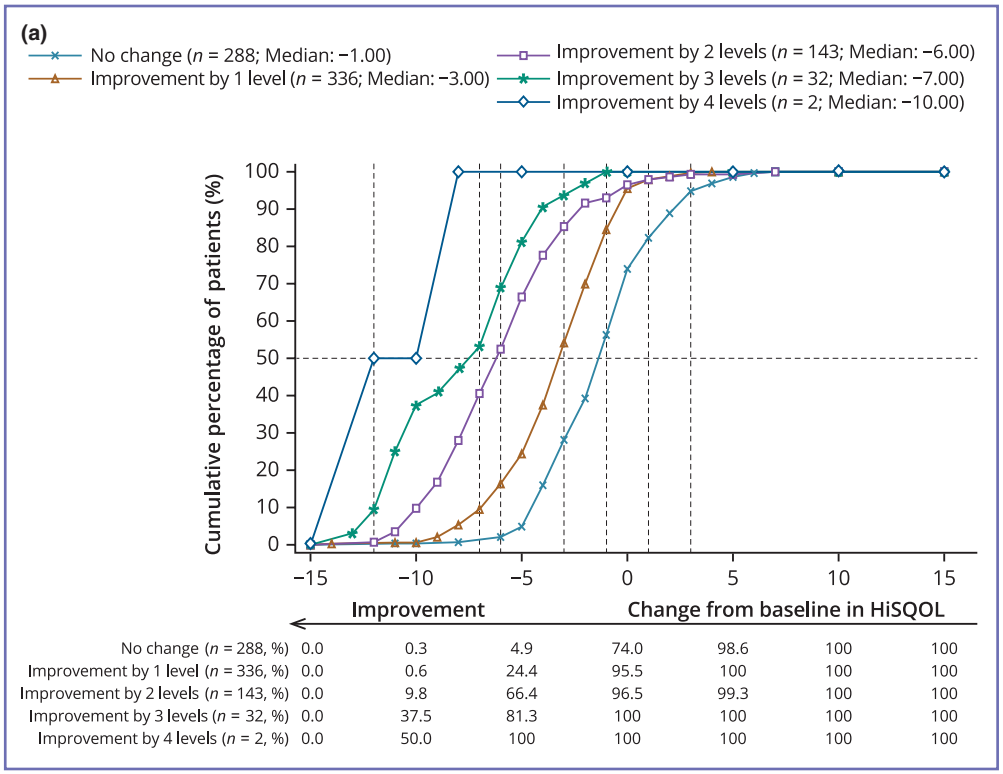
CFA models, with adequate loadings (> 0.40) onto their hypothesized group factor, even though model fit was suboptimal for the three- and one-factor models, supported the relevance of the three HiSQOL subscales. Bi-factor and one-factor models supported the relevance of an underlying unidimensional concept from all HiSQOL items, and thus a total score derived directly from HiSQOL items. These results are generally consistent with published findings from an observational study conducted in the USA and Denmark.[11]

The HiSQOL demonstrated construct validity, with all convergent correlations in the expected positive direction and strength, moderate-to-strong, at baseline and week 16 for DLQI total score. At baseline, HiSQOL subscale scores and the total score were only weakly correlated with the IHS4, while hypothesized to be moderately-to-strongly correlated. At week 16, correlations were moderate, except for the psychosocial subscale, which was weak. The lower correlations between the HiSQOL and IHS4 are likely to be due to the concepts captured by each measure being slightly different, in addition to the divergence between clinician-reported outcomes and patient-reported outcomes that has been reported in the literature across disease areas.[29–32]

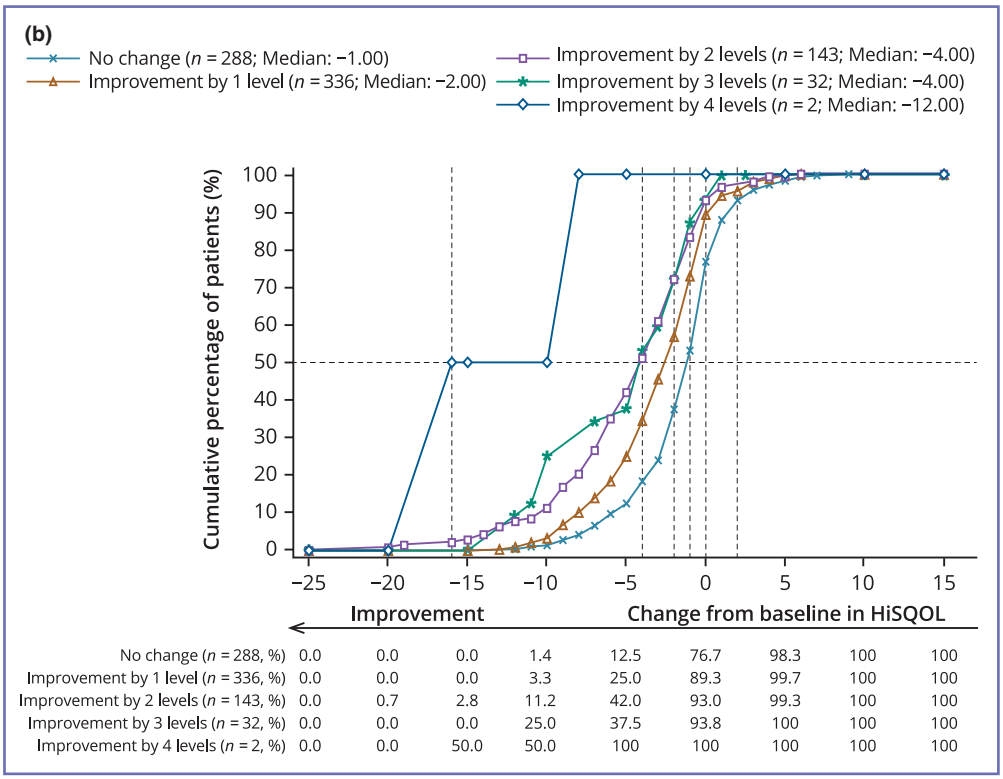
Known-groups analyses generally yielded results as expected. Mean HiSQOL subscale and total scores in groups with more severe Hurley stage, PGI-S-HS, IHS4 and HS PGA were generally higher.

High internal consistency reliability was observed across subscale and total scores at baseline and week 16. Acceptable test–retest reliability was observed in all cases between baseline and weeks 4, 32 and 48. Additionally, correlation analyses between changes from baseline to week 16 in HiSQOL score against changes in PGI-S-HS indicated that HiSQOL subscale and total scores had sufficient sensitivity to detect changes in measures assessing a similar construct.

From anchor-based and distribution-based analyses, the following ranges are recommended to represent clinically meaningful within-patient improvements: symptoms subscale score, 5–6-point decrease; psychosocial subscale score, 4–5-point decrease; activities–adaptations subscale score, 10–11-point decrease; and total score, 20–21-point decrease. Thresholds were derived using a 2-point reduction on the 5-point PGI-S-HS (moving from a very severe to a moderate state, or from a severe to a mild state), to represent clinically meaningful improvements, as recommended by the US Food and Drug Administration in the

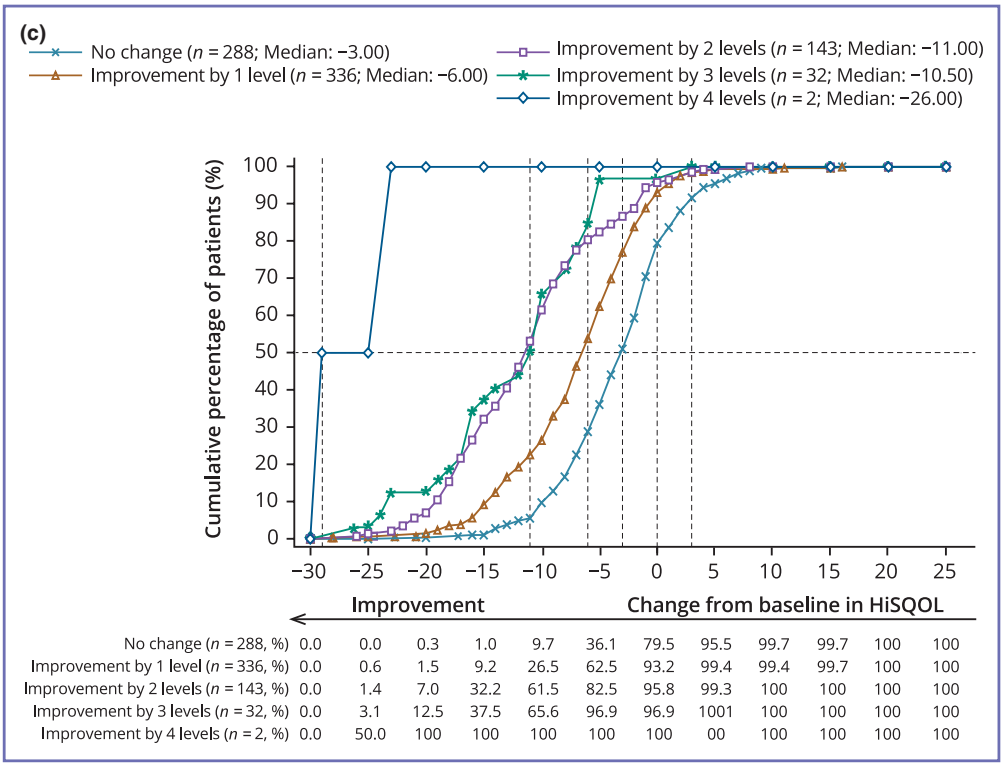


No change (*n* = 288, %)	0.0	0.3	4.9	74.0	98.6	100	100
Improvement by 1 level (*n* = 336, %)	0.0	0.6	24.4	95.5	100	100	100
Improvement by 2 levels (*n* = 143, %)	0.0	9.8	66.4	96.5	99.3	100	100
Improvement by 3 levels (*n* = 32, %)	0.0	37.5	81.3	100	100	100	100
Improvement by 4 levels (*n* = 2, %)	0.0	50.0	100	100	100	100	100

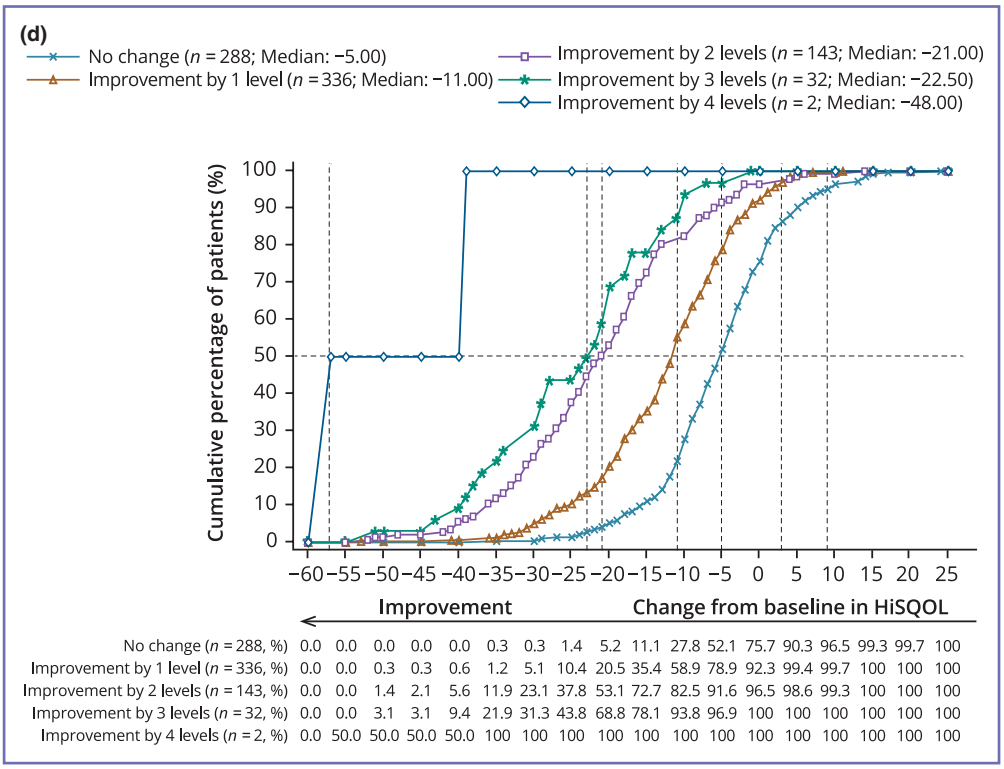


No change (*n* = 288, %)	0.0	0.0	0.0	1.4	12.5	76.7	98.3	100	100
Improvement by 1 level (*n* = 336, %)	0.0	0.0	0.0	3.3	25.0	89.3	99.7	100	100
Improvement by 2 levels (*n* = 143, %)	0.0	0.7	2.8	11.2	42.0	93.0	99.3	100	100
Improvement by 3 levels (*n* = 32, %)	0.0	0.0	0.0	25.0	37.5	93.8	100	100	100
Improvement by 4 levels (*n* = 2, %)	0.0	0.0	50.0	50.0	100	100	100	100	100

Figure 2 Empirical cumulative distribution function curves of changes from baseline to week 16 on (a) HiSQOL symptoms score; (b) HiSQOL psychosocial score; (c) HiSQOL activities–adaptations score and (d) HiSQOL total score, by change in levels of PGI-S-HS response category (none, mild, moderate, severe, very severe).HiSQOL, Hidradenitis Suppurativa Quality of Life questionnaire; HS: hidradenitis suppurativa; PGI-S-HS, patient-global impression of HS severity.



No change (*n* = 288, %)	0.0	0.0	0.3	1.0	9.7	36.1	79.5	95.5	99.7	99.7	100	100
Improvement by 1 level (*n* = 336, %)	0.0	0.6	1.5	9.2	26.5	62.5	93.2	99.4	99.4	99.7	100	100
Improvement by 2 levels (*n* = 143, %)	0.0	1.4	7.0	32.2	61.5	82.5	95.8	99.3	100	100	100	100
Improvement by 3 levels (*n* = 32, %)	0.0	3.1	12.5	37.5	65.6	96.9	96.9	1001	100	100	100	100
Improvement by 4 levels (*n* = 2, %)	0.0	50.0	100	100	100	100	100	00	100	100	100	100



No change (*n* = 288, %)	0.0	0.0	0.0	0.0	0.0	0.3	0.3	1.4	5.2	11.1	27.8	52.1	75.7	90.3	96.5	99.3	99.7	100
Improvement by 1 level (*n* = 336, %)	0.0	0.0	0.3	0.3	0.6	1.2	5.1	10.4	20.5	35.4	58.9	78.9	92.3	99.4	99.7	100	100	100
Improvement by 2 levels (*n* = 143, %)	0.0	0.0	1.4	2.1	5.6	11.9	23.1	37.8	53.1	72.7	82.5	91.6	96.5	98.6	99.3	100	100	100
Improvement by 3 levels (*n* = 32, %)	0.0	0.0	3.1	3.1	9.4	21.9	31.3	43.8	68.8	78.1	93.8	96.9	100	100	100	100	100	100
Improvement by 4 levels (*n* = 2, %)	0.0	50.0	50.0	50.0	50.0	100	100	100	100	100	100	100	100	100	100	100	100	100

Figure 2 (*Continued*)

Table 5 Severity thresholds for HiSQOL

	None	Mild	Moderate	Severe	Very severe
HiSQOL symptoms score	0–2	3–4	5–6	7	8–16
HiSQOL psychosocial score	0	1	2	3–5	6–20
HiSQOL activities–adaptations score	0–1	2–4	5–8	9–12	13–32
HiSQOL total score	0–4	5–14	15–21	22–23	24–68

HiSQOL, Hidradenitis Suppurativa Quality of Life questionnaire.

Patient-Focused Drug Development guidance documents. Using a 1-point reduction to derive the thresholds would rather lead to derivation of a minimally clinically important difference (MCID). Also referred to as minimal important difference, MCID was first introduced by Jaeschke *et al.* as 'the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management'.[33] In the present study, those thresholds are around half the clinically meaningful within-patient improvement thresholds and represent approximately 15% of the score range. The latter aligns with published estimates of the DLQI MCID, which varied from 3 to 5 on a 0–30-point scale, representing 10–15% of the score range.[34] Severity thresholds were also estimated with good discriminant power.

By focusing on symptoms, activities–adaptations and psychosocial consequences that impact daily activities and wellbeing, the HiSQOL provides a clinically meaningful and holistic view of the disease beyond the lesions to patients, treating physicians and patient organizations. The utility of the HiSQOL in different settings has been demonstrated, as HiSTORIC recently released their recommendation of HiSQOL to assess patients' HRQoL in routine clinical practice.[35]

There are some limitations. The use of data from phase III trial populations required to meet certain eligibility criteria skewed the sample to adult patients with moderate-to-severe HS and may limit broader extrapolation, particularly to those with mild disease. Conversely, the multinational nature of BE HEARD and the use of HiSQOL translations is a strength because HiSQOL validation can be considered in different languages, although no cross-country comparisons were made. This study was not able to determine clinically meaningful within-patient worsening thresholds, due to small sample sizes. This was anticipated considering the context of the clinical trial, as most patients were treated with active treatment and were thus expected to improve. Future studies are needed to confirm the psychometric properties and thresholds in contexts other than clinical trials.

In conclusion, this was the first time the HiSQOL was prospectively used in a phase III development programme to capture the experience of patients with HS, presenting an opportunity to understand its psychometric properties and derive interpretation thresholds. Overall, results demonstrated robust properties of HiSQOL subscale and total scores in a substantial, representative sample of patients with moderate-to-severe HS. These results provide clinically meaningful differences and score severity bands for the HiSQOL total and subscale scores, supporting interpretation of trial results and informing treatment decisions.

Acknowledgements

The authors thank the patients, the investigators and their teams who took part in this study. The authors also acknowledge Luke Peterson, UCB, Morrisville, USA, for statistical analysis support; Susanne Wiegratz MSc, UCB, Monheim am Rhein, Germany, for publication coordination and Charlotte Marris PhD, from Costello Medical, UK, for medical writing and editorial assistance based on the authors' input and direction.

Funding sources

This study was sponsored by UCB. Support for third-party writing assistance for this article, provided by Charlotte Marris PhD, Costello Medical, UK, was funded by UCB in accordance with Good Publication Practice (GPP) 2022 guidelines (https://www.ismpp.org/gpp-2022). UCB contributed to study design, participated in data collection, completed the data analysis and participated in data interpretation. UCB participated in the writing, review and approval of the manuscript. All authors had full access to the data, reviewed and approved the final version and were responsible for the decision to submit for publication. A medical writing agency, employed by UCB, assisted with manuscript preparation under the authors' direction.

Conflicts of interest

J.S.K. reports personal fees from AbbVie, ChemoCentryx, CSL Behring, DermTech, Incyte, Insmed, Janssen, MoonLake Immunotherapeutics, Novartis and UCB; personal fees and grants from Incyte; is a co-copyright holder of HiSQOL©; is a consultant for and has received honoraria from AbbVie, Alumis, DermTech, Incyte, Insmed, Janssen, MoonLake Immunotherapeutics, Novartis and UCB; and has been a speaker for AbbVie, Janssen, Novartis and UCB. L.T. received speaker honoraria from UCB; and is co-copyright holder of HiSQOL© and HS-IGA. J.L., V.C., R.R., I.P. and E.M. are employees and shareholders of UCB. C.G.P. is an employee of Evidera, a part of Thermo Fisher Scientific that receives funding for research from UCB. J.R.I. receives an authorship honorarium from UpToDate; is a consultant for AbbVie, Boehringer Ingelheim, ChemoCentryx, Citryll, MoonLake Immunotherapuetics, Novartis, UCB and Union Therapeutics and has served on advisory boards for Insmed, Kymera Therapeutics and Viela Bio; is a co-copyright holder of HiSQOL© and HS-IGA; his department receives income from copyright of the Dermatology Life Quality Instrument (DLQI) and related instruments.

Data availability

Underlying data from this manuscript may be requested by qualified researchers 6 months after product approval in the USA and/or Europe, or global development is discontinued, and 18 months after trial completion. Investigators may request access to anonymized individual patient-level data and redacted trial documents which may include analysis-ready datasets, study protocol, annotated case report form, statistical analysis plan, dataset specifications and clinical study report. Prior to use of the data, proposals need to be approved by an independent review panel at www.Vivli.org and a signed data-sharing agreement will need to be executed. All documents are available in English only, for a prespecified time, typically 12 months, on a password-protected portal.

Ethics statement

The study protocol, amendments and patient-informed consent were reviewed by a national, regional or independent ethics committee (IEC) or institutional review board (IRB). This study was conducted in accordance with the current version of the applicable regulatory and International Conference on Harmonisation (ICH)-Good Clinical Practice (GCP) requirements, the ethical principles that have their origin in the principles of the Declaration of Helsinki, and the local laws of the countries involved.

Patient consent

Not applicable.

Trial registration

BE HEARD I (NCT04242446) and II (NCT04242498).

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website.

References

1 Sabat R, Jemec GBE, Matusiak Ł *et al.* Hidradenitis suppurativa. *Nat Rev Dis Primers* 2020; **6**:18.
2 Matusiak L. Profound consequences of hidradenitis suppurativa: a review. *Br J Dermatol* 2020; **183**:e171–7.
3 Garg A, Neuren E, Cha D *et al.* Evaluating patients' unmet needs in hidradenitis suppurativa: results from the Global Survey Of Impact and Healthcare Needs (VOICE) Project. *J Am Acad Dermatol* 2020; **82**:366–76.
4 Wright S, Strunk A, Garg A. New-onset depression among children, adolescents, and adults with hidradenitis suppurativa. *J Am Acad Dermatol* 2020; **83**:1360–6.
5 Wright S, Strunk A, Garg A. Prevalence of depression among children, adolescents, and adults with hidradenitis suppurativa. *J Am Acad Dermatol* 2022; **86**:55–60.
6 Zouboulis CC, Gulliver W, Ingram J *et al.* Endpoints of clinical trials for hidradenitis suppurativa: proceedings of a round-table session. *Exp Dermatol* 2020; **29** (Suppl. 1):67–72.
7 Thorlacius L, Ingram JR, Villumsen B *et al.* A core domain set for hidradenitis suppurativa trial outcomes: an international Delphi process. *Br J Dermatol* 2018; **179**:642–50.
8 Finlay AY, Khan G. Dermatology Life Quality Index (DLQI) – a simple practical measure for routine clinical use. *Clin Exp Dermatol* 1994; **19**:210–16.
9 Thorlacius L, Esmann S, Miller I *et al.* Development of HiSQOL: a hidradenitis suppurativa-specific quality of life instrument. *Skin Appendage Disord* 2019; **5**:221–9.
10 Chren M-M, Lasek RJ, Flocke SA, Zyzanski SJ. Improved discriminative and evaluative capability of a refined version of Skindex, a quality-of-life instrument for patients with skin diseases. *Arch Dermatol* 1997; **133**:1433–40.
11 Kirby J, Thorlacius L, Villumsen B *et al.* The Hidradenitis Suppurativa Quality of Life (HiSQOL) score: development and validation of a measure for clinical trials. *Br J Dermatol* 2020; **183**:340–8.
12 US Food and Drug Administration. Patient-focused drug development: methods to identify what is important to patients. Guidance for industry, Food and Drug Administration staff, and other stakeholders. February 2022. Available from: https://www.fda.gov/media/131230/download.
13 US Food and Drug Administration. Patient-focused drug development: incorporating clinical outcome assessments into endpoints for regulatory decision-making. Guidance for industry, Food and Drug Administration staff, and other stakeholders. Draft guidance. April 2023. Available from: https://www.fda.gov/media/166830/download.
14 US Food and Drug Administration. Guidance for Industry: patient-reported outcome measures: use in medical product development to support labeling claims. December 2009. Available from: https://www.fda.gov/media/77832/download.
15 ClinicalTrials.gov. A study to evaluate the efficacy and safety of bimekizumab in study participants with moderate to severe hidradenitis suppurativa (BE HEARD I). Available from: https://clinicaltrials.gov/study/NCT04242446.
16 ClinicalTrials.gov. A study to evaluate the efficacy and safety of bimekizumab in study participants with moderate to severe hidradenitis suppurativa (BE HEARD II). Available from: https://clinicaltrials.gov/study/NCT04242498.
17 Kimball AB, Jemec GBE, Sayed CJ *et al.* Efficacy and safety of bimekizumab in patients with moderate-to-severe hidradenitis suppurativa (BE HEARD I and BE HEARD II): two 48-week, randomised, double-blind, placebo-controlled, multicentre phase 3 trials. *Lancet* 2024; **403**:2504–19.
18 Hurley HJ. Axillary hyperhidrosis, apocrine bromhidrosis, hidradenitis suppurativa, and familial benign pemphigus: surgical approach. In: *Dermatologic Surgery* (Roenigk RK, Roenigk HH, eds). New York: Marcel Dekker, 1989; 729–39.
19 Cappelleri JC, Zou KH, Bushmakin AG *et al. Patient-Reported Outcomes: Measurement, implementation and interpretation*. CRC Press (Taylor & Francis Group), 2014.
20 Bryne B. *Structural Equation Modeling with EQS: Basic concepts, applications, and programming*, 2nd edn. Lawrence Erlbaum Associates Inc., 2006.
21 O'Rourke N, Hatcher L, Stepanski E. *A Step-by-Step Approach to Using SAS for Univariate & Multivariate Statistics*. SAS Institute, 2005.
22 Patel ZS, Hoffman LK, Buse DC *et al.* Pain, psychological comorbidities, disability, and impaired quality of life in hidradenitis suppurativa. *Curr Pain Headache Rep* 2017; **21**:49 [correction published 18 November 2017]. https://doi.org/10.1007/s11916-017-0647-3.
23 Cook KF, Kallen MA, Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Quality Life Res* 2009; **18**:447–60.

24 Hinkle DE, Wiersma W, Jurs SG. *Applied Statistics for the Behavioral Sciences*, 5th edn. Houghton Mifflin, 2003.
25 Aaronson N, Alonso J, Burnam AL *et al.* Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual of life Res* 2002; **11**:193–205.
26 Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD: J Chronic Obstruct Pulmonary Dis* 2005; **2**:63–7.
27 Hosmer DW, Lemeshow S. *Applied Logistic Regression*, 2nd edn. New York: Wiley; 2000. xii, 373.
28 Centers for Disease Control and Prevention (CDC). Adult BMI Categories. 19 March 2024. Available from: https://www.cdc.gov/bmi/adult-calculator/bmi-categories.html.
29 Barbieri JS, Gelfand JM. Patient-reported outcome measures as complementary information to clinician-reported outcome measures in patients with psoriasis. *JAMA Dermatol* 2021; **157**:1236–7.
30 Bily W, Jauker J, Nics H *et al.* Associations between patient-reported and clinician-reported outcome measures in patients after traumatic injuries of the lower limb. *Int J Environ Res Public Health* 2022; **19**:3140.
31 Long V, Chen Z, Du R *et al.* Understanding discordant perceptions of disease severity between physicians and patients with eczema and psoriasis using structural equation modeling. *JAMA Dermatol* 2023; **159**:811–19.
32 Xiao C, Polomano R, Bruner DW. Comparison between patient-reported and clinician-observed symptoms in oncology. *Cancer Nurs* 2013; **36**:E1–16.
33 Jaeschke R, Singer J, Guyatt G. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; **10**:407–15.
34 Basra M, Salek M, Camilleri L *et al.* Determining the minimal clinically important difference and responsiveness of the Dermatology Life Quality Index (DLQI): further data. *Dermatology* 2015; **230**:27–33.
35 Mastacouris N, Tannenbaum R, Strunk A *et al.* Outcome measures for the evaluation of treatment response in hidradenitis suppurativa for clinical practice: a HiSTORIC consensus statement. *JAMA Dermatol* 2023; **159**:1258–66.