The Hidradenitis Suppurativa Quality of Life (HiSQOL) score: Development and validation of a measure for clinical trials

Joslyn S Kirby, MD, MS, MEd[1] *, Linnea Thorlacius ,MD, PhD[2] *, Bente Villumsen[3], John R Ingram[4], Amit Garg[5], Karl Bang Christensen[2], Melissa Butt, MPH[1], Solveig Esmann[2], Jerry Tan, MD[6]**, Gregor BE Jemec, MD[2]**

*co-first authors
**co-senior authors

[1] Department of Dermatology, Penn State Health Milton S. Hershey Medical Center, Hershey, PA, USA
[2] Department of Dermatology, Zealand University Hospital, Roskilde; Health Sciences Faculty, University of Copenhagen, Denmark
[3] The Patients' Association HS Denmark, Copenhagen, Denmark.
[4] Division of Infection & Immunity, Cardiff University, University Hospital of Wales, Heath Park, Cardiff, CF14 4XN, United Kingdom
[5] Department of Dermatology, Donald and Barbara Zucker School of Medicine at Hofstra Northwell, New Hyde Park, New York
[6] Department of Medicine, University of Western Ontario, Windsor, Ontario, Canada

Corresponding author:
Joslyn S Kirby, MD, MS, MEd
Associate Professor
Penn State Milton S Hershey Medical Center
500 University Dr, HU14
Hershey PA 17033
Jkirby1@pennstatehealth.psu.edu
1-717-531-8307

Funding Sources:
Dr Kirby received funding from the Agency for Healthcare Research and Quality for this research (K08HS024585).
Use of REDCap through Penn State is supported by NIH/NCATS Grant Number UL1 TR000127 and UL1 TR002014 through The Penn State Clinical & Translational Research Institute, Pennsylvania State University CTSA
Dr Ingram received funding from Health and Care Research Wales (Health Fellowship 14-08).

IRB approval status: IRB approved

Conflicts of Interest:
Kirby: AbbVie: Speaker, Advisory Board (Honoraria), Investigator; Incyte, ChemoCentryx: Consultant (Fees), Investigator; UCB: Investigator; InflaRx: Investigator
Thorlacius: Abbvie, Janssen: travel expenses. Regeneron: Investigator
Garg: Advisor for AbbVie, Pfizer, Janssen, Asana Biosciences, and UCB (honoraria)
Ingram: UCB Pharma, Novartis: Consultant; Abbvie: travel expenses.
Tan: UCB Advisory Board (Honoraria); Incyte: Investigator
Jemec: Advisory Board (honoraria): AbbVie, Chemocentryx, Coloplast, Incyte, Inflarx, Novartis, Pierre Fabre and UCB; Abbvie, Leo Pharma, Janssen-Cilag, Regeneron, Sanofi, Astra-Zeneca and Novartis: Investigator; AbbVie, Boehringer-Ingelheim, Galderma and MSD: speaker (honoraria); Abbvie, Leo Pharma and Novartis: unrestricted grants.
Villumsen, Christensen, Butt, Esmann: None

Key point word count: 96/140
Abstract word count: 239/250
Manuscript word count: 3,419/3000
References: 59
Figures: 1
Tables: 3
Supplements: 5

Acknowledgment statement: The authors would like to thank and acknowledge the individuals with hidradenitis suppurativa who took part in this study and whose time and effort made this study possible including Angela Gibbons, Paul Gorman, Tiffany Mojica, Alison Wright and others.

Key Questions

What is already known about this topic?

- HS is a chronic, relapsing inflammatory skin condition with potential adverse impacts on health-related quality of life.
- The ability to assess HS-specific HRQOL is important to those with HS and to furthering research to mitigate the effects of the condition.
- Development of HS-specific instruments is feasible and existing instruments have limitations.

What does this study add?

- This study describes the development, validation, and psychometric properties of the HiSQOL, a novel HS-specific HRQOL instrument.
- HiSQOL is a patient-reported outcome measure developed for clinical trials to address disease-specific changes in HRQOL.

Abstract

Background: Hidradenitis suppurativa (HS) is a chronic, inflammatory condition that can have a large negative impact on health-related quality of life (HRQOL). A reliable and validated measure of HS-specific HRQOL in clinical studies is needed.

Objective: To develop and validate the Hidradenitis Suppurtiva Quality Of Life (HiSQOL©) scale, for clinical trial measurement of HS-specific HRQOL.

Methods: Stage 1: Qualitative concept elicitation (CE) interviews were conducted with HS patients in Denmark (DK) (n = 21) and the United States (US) (n=21). Stage 2: Cognitive debriefing (CD) interviews were performed with US HS patients (n = 30) and Danish HS patients (n=30). Stage 3: Observational study of 222 HS patients in the US was conducted for item reduction, measure validation and assessment of psychometric properties. Stage 4: Observational study of 215 HS patients in Denmark was conducted to confirm the psychometric structure derived in stage 3. In both studies - the Dermatology Life Quality Index, Hospital Anxiety and Depression Scale, and numerical rating scale for pain - were also included.

Results: In CE, 99 items were generated and reduced to 41 after removing duplicates. In CD, 2 items were added and 1 items removed. A 42-item instrument was psychometrically assessed. Based on psychometric analyses and patient input, the instrument was reduced to 17 items that had strong psychometric properties in both US and DK samples.

Discussion: The HiSQOL is a reliable and valid instrument to measure HS-specific HRQOL for clinical trials.

Introduction

Hidradenitis suppurativa (HS) is a chronic, recurring inflammatory skin condition most commonly affecting the groin, axillae, buttocks, and inframammary folds[1]. Clinical manifestations include inflamed nodules, abscesses, tunnels that cause pain, itch, drainage, odor and often eventuate into scars or post-inflammatory hyperpigmentation[2-4]. HS can make activities of daily living, such as walking, sitting, and working difficult or impossible. HS lesions and the malodorous drainage can be socially and emotionally devastating[5,6]. Thus, HS has repeatedly been shown to have a large negative impact on health-related quality of life (HRQOL) [7-10], severe socio-economic consequences[11,12] and even an increased risk of suicide[13].

In an effort to the enhance quality and consistency among future treatment studies, a core outcome set was recently established for HS[14-16]. The HIdradenitis Suppurativa cORe outcomes set International Collaboration (HISTORIC) reached consensus on a core outcome set that specifically recommended assessment of HS-specific HRQOL. HRQOL scales include constructs that are of primary concern to the patient and may include the impact of symptoms, functional impairments and emotions on HRQOL[17,18]. Several generic dermatologic HRQOL scales exist, however, generic HRQOL measures do not assess the unique and important ways that HS affects patients due to the distinctive symptoms and location of the condition[9,19,20]. Thus, the HS core outcome set established the need for an HS-specific HRQOL that included: physical functioning, psychological functioning, psychosocial functioning, emotional well-being, and ability to work or study[14]. There has been a burst of activity to address the need for a HS-specific HRQOL instrument and several instruments, as reviewed by Chernyshov et al[10], have been recently developed. Each instrument has strengths and limitations, such as limited evidence of validity and/or reliability or prolonged recall period[10,21-27]. Also, the focus of the core outcome set is to ensure rigorous measurement in interventional clinical trials, so some constructs such as scarring, skin damage, or body image may be less amenable to change in the shorter time-frame of clinical trials. To address the need for a rigorously-developed and psychometrically-sound HS-specific HRQOL instrument, this group sought to develop and test the Hidradenitis Suppurativa Quality of Life (HiSQOL) tool, an instrument designed to measure HS-specific HRQOL of adults with HS in the setting of a clinical trial.

Methods

Study design and Participants

A mixed methods design was utilized and included four phases aligned with guidance from the US Food and Drug Administration[28]. Conduct of the study was overseen by an international group of investigators, which included patient research partners, clinicians with expertise in HS, and researchers with expertise in instrument development. People with HS who were 18 years or older were identified based on diagnostic code for HS (International Classification of Diseases, Ninth Revision (ICD-9) code 705.83) in the medical record at two academic institutions in the United States (US) and Demark (DK). People who gave informed consent, had a confirmed diagnosis of HS, and were fluent in English or Danish were recruited by phone and in clinic. This study was approved by the ethics committee of each institution and the Danish Data Protection Agency. All participants gave written informed consent prior to participation in the study.

Concept Elicitation (CE)

Semi-structured interviews were conducted with people with HS by two investigators with experience (JS, ES). Interviews included open-ended questions with follow-up probing questions (Interview guide in Supplement 1). Interviews were audiotaped and transcribed verbatim. Qualitative analysis was conducted independently by two researchers in the original language by a native speaker with Nvivo 11 software (QSR International, Burlington, MA). Using grounded theory methods[29] and qualitative analysis software, quotations were assigned a code determined by the underlying concept and grouped into higher level concepts. Coding was informed by the model of HRQOL by Ferrans et al[30] as it was shown to better explain HRQOL[31]. Conceptual saturation was assessed and achieved.

Instrument Development

Items were developed using the qualitative data and the HRQOL model[30] by clinicians with expertise in HS and four patient research partners were present to ensure the items were relevant and comprehensive. Concepts related to active disease were included and concepts clearly related only to secondary skin damage, e.g. scarring, were excluded since the anticipated use of the tool is a clinical trial setting where changes in active disease but not secondary damage are anticipated. Through discussion, the group condensed or eliminated duplicate data. The extant literature guided design of the recall period, item stems, response scale, and instructions[32]. The initial instrument was translated into Danish based on recognized methods for cross-cultural adaptation[33]. Briefly, two bilingual translators whose first language was Danish produced two independent translations. One translator was aware of the concepts being examined in the instrument, the other was not. An observer synthesized a single common form for back-translation, and then two native English speakers without a medical background independently translated the form. These forms were consolidated by a committee

of methodologists, health professionals, and the translators.

Cognitive Debriefing (CD) / Pilot-testing

Interviews and focus groups were conducted to evaluate the relevance of the concepts evaluated by the items (content validity), the ability of the target audience (English- or Danish-fluent adults with HS) to understand and complete the instrument, completeness, and acceptability (Interview guide in Supplement 2). Per CE methods, people with HS were recruited in the US and DK and excluded prior CE participants. Participants were asked to complete the instrument using the "think-aloud" technique, which facilitates feedback on the instrument. The interviewers (JSK, ES) also asked probing questions to elicit suggestions. The combined use of these is a rigorous approach to establish whether respondents understand the questions in the way the researcher intended[34,35]. As per CE, analysis was conducted with Nvivo 11 software (QSR International, Burlington, MA).

Field-testing and Psychometric Assessment

An observational non-interventional non-randomized study was conducted in the US and DK for field-testing and further psychometric validation of the HiSQOL candidate items. The field-testing aimed for item reduction, examination of dimensionality and definitive selection of items per dimension. Per CE, eligibility criteria were applied to identify participants. The HiSQOL instrument, Dermatology Life Quality Index (DLQI)[36,37], the Hospital Anxiety and Depression Scale (HADS)[38,39], and numerical rating scale (NRS) for pain[40] were administered concurrently. A web version of all instruments and items was developed in REDCap (Research Electronic Data Capture), a secure, web-based application designed to support data capture for research studies[41]. The sample was divided into a development sample for item reduction and initial analyses (US sample), and a validation sample (DK sample). HiSQOL and a patient-rated perception of change in HS item were administered a second time 24-72 hours later to evaluate test-retest reliability. This timeframe was chosen due to the unpredictable, intermittent, and rapid onset of HS worsening.

Analysis

Item response distributions, inter-item correlations, item-total correlations, as well as multiple aspects of reliability and validity were evaluated for the candidate instrument using complete responses. Confirmatory factor analysis (CFA) and item response theory (IRT), with the graphical log linear Rasch model were used to evaluate the items in the long form[42,43]. Item fit was evaluated by dividing the total score into class intervals and plotting observed item means against score intervals together with 99% confidence bands. Differential item functioning (DIF) was evaluated using Mantel-

Haentzel test, while local dependence (LD) was evaluated using Yens Q3[44]. When a value was more than 0.2 above the average residual correlation it was considered evidence of LD, i.e. when Q3,* was larger than 0.2. For all CFA models DIF was added by allowing item thresholds to be different across gender or age group and LD was added by including correlated error terms. Unfavorable response distributions, inter-item correlation, IRT and CFA data, and DIF or LD results were taken into account in the item reduction process, which was overseen by the investigators and three patient research partners. We used the US sample as a calibration set to identify a shortened instrument, then did a preliminary validation of short form with the US sample. The DK sample was used to confirm the short form.

The sub-scale structure was investigated by comparing a three-dimensional to a bifactor CFA model. It was hypothesized that the bifactor model would fit the data better indicating that an overall HiSQOL score can be reported alongside domain scores. All CFA models were fitted using M Plus 6th edition (Muthen & Muthen, Los Angeles). Fit of the CFA was evaluated based on the Chi-square test of model fit, the root mean square error of approximation (RMSEA), the comparative fit index (CFI) and the Tucker Lewis Index (TLI). For the RMSEA, a smaller value indicates a closer fit; an RMSEA <.06 is considered to reflect good fit, values <.08 are fair, and values above .10 are generally considered to reflect poor fit. Values of the CFI and the TLI above .95 are generally accepted as reflecting adequate and good fit.

For convergent validity, it was hypothesized that there would be at least moderate correlation between the scores of the HiSQOL instrument and DLQI, HADS, and NRS for pain. This relationship was assessed using Spearman's rank-sum correlations. A correlation of 0–0.09 was considered no correlation, 0.1≤0.3 was considered poor, 0.31≤0.6 was considered fair, 0.61≤0.8 was moderate and 0.81<1 was considered very strong, equal to 1 was considered perfect[45]. Known groups validity was evaluated as the differences in HiSQOL scores among known scoring bands for the DLQI using ANOVA[46,47]. It was hypothesized there would be a significant difference in the mean HiSQOL score among the DLQI known score bands. Cronbach's alpha was used to evaluate internal consistency reliability of the instrument. Test-retest reliability was assessed with two instances of complete data in the US sample and was assessed using intra-class correlations. Participants who reported stable HS were included in this analysis. The standard sample size for convergent validity calculation is a minimum of five subjects per item[48-50], so a sample size of 225 was estimated based on the 45-item scale. This sample size was adequate for test-retest reliability calculations, per guidelines of Bland and Altman[51], so with 2 repetitions and requiring within-subject standard deviation be within 10% of the population value, the minimum sample size was 192.

Results

Demographic and clinical characteristics of participants in CE and CD stages

Table 1 details the demographic and clinical characteristics for the CE and CD samples. Race and ethnicity information was collected with US participants, but was not collected with Danish participants per protocol.

Content elicitation

Participants most frequently discussed the impact of symptoms as well as psychosocial effects and alterations in functions and activities. Examples of patient quotations and the major themes/concepts are available in Supplemental Table 3. Saturation was achieved within both country samples. No country-level differences in the main concepts or sub-concepts were noted. As a result, a conceptual framework was developed and used to generate items to measure the core HRQOL impacts due to HS.

Instrument development

Based on the CE data and extant literature[14,30,52], the investigators generated 99 items. Concepts related to active disease were included and concepts clearly related skin damage were excluded since the anticipated use of the instrument is a clinical trial setting where changes in active disease but not damage are expected. The research team including four patient research partners iteratively discussed the item meaning and condensed or eliminated duplicate data, then grouped items with one concept of the conceptual model. A 7-day recall period was chosen to capture short-term changes. [32] A 5-point item response scale incorporating “extremely,” “very much,” “moderately,” “slightly,” or “not at all” was used for all items. The point value assigned to these responses was 4, 3, 2, 1, and 0, respectively. For some items, respondents were given the additional option of “Unable to do, due to my HS” and/or “I do not normally do this, HS did not influence.” The former option was assigned a score of 4 to indicate the severity of the impact of HS, whereas the latter option was assigned a score of 0 to indicate HS did not impact it.

Cognitive debriefing / Pilot testing

Two phases of cognitive debriefing (CD) interviews and focus groups were conducted. The second round was conducted to ensure that changes made after the first round were acceptable. Participant characteristics are presented in Table 1. Participants indicated that the instrument assessed relevant symptoms and impacts. Patients did not identify any missing items. In CD phase 1, instructions and items were reorganized or wording simplified. The ‘Concentration Consequences’ section included only one item so two items were added to more robustly evaluate this construct. In CD phase 2, minor

wording changes were made to item responses and no items were added. One item was removed because it was felt to represent a global HRQOL question, resulting in a 42-item instrument.

Psychometric Assessment / Field testing

Table 1 lists the characteristics of the eligible patients included in this stage. Forty-seven completed instruments were excluded as the participants did not meet the inclusion criteria. Most participants were female, Caucasian, and Hurley stage II; however, there was participation across a range of respondents including males, Black, Hispanic, and Hurley stage I and stage III participants.

Item reduction was conducted with the aim of retaining the most discriminative items and at least one item for each concept in the conceptual framework as well the core outcome set. Results of analyses along with input from the study team and five people with HS were used to identify a shortened 17-item instrument that maintained content coverage with maximum precision. Twenty-eight items were deleted due to: floor/ceiling effects, lack of applicability to most people with HS due to specificity of the item for a body site, IRT item fit, or DIF with respect to sex. The 17-item HiSQOL included four symptom items, eight activity-adaptation items, and five psychosocial items. The item scores are summed to create a total ranging from 0 to 68, with higher scores indicating more severe impact on HRQOL. The sub-scale scores range from 0 to 16 for symptoms, 0 to 20 for psychosocial, and 0 to 32 for activities-adaptations.

For the symptoms subscale there was evidence of LD for the item pair ‘Pain’ and ‘Itch’ in both samples (Q3,*=0.26), while evidence of gender DIF for the item ‘Itch’ was found in the Danish sample only. In the multiple groups CFA there was no evidence of DIF, but the item ‘Drainage’ functioned differentially across the two samples. For the psychosocial subscale there was evidence of LD for the item pair 'Anxious or nervous' and ‘Concentration’ in both samples (Q3,*=0.20) and for the item pair 'Embarrassed' and ‘Sexual desire’ in the Danish sample only. Regarding DIF there was evidence of gender DIF for the item ‘Concentration’ in the US sample only. For the Activities-adaptations subscale there was evidence of LD for the item pair ‘Washing yourself’ and ‘Getting dressed’ in both samples (Q3,*=0.46) and for the item pair ‘Walking’ and ‘Exercising’ in US sample only. There was evidence of gender DIF for the item ‘What you wear’ in the US sample.

Table 2 shows descriptive statistics, range of inter-item correlations and the item-total correlations of the psychometric evaluation using the US sample followed by validation using the Danish sample. Structural construct validity was established by CFA that confirmed fit of a bifactor model (Chi-Square=633.1, df=342, P<0.0001, RMSEA=0.062 (90% CI 0.055 to 0.070), CFI=0.978, TLI=0.976) indicating that the total HiSQOL score or sub-scale scores can be utilized in assessment. The bifactor model fitted the data better than a three-dimensional CFA model. The model derived for

the three subscales for symptoms, psychosocial, and activities-adaptations using multiple groups CFA all showed excellent fit to the data (Supplement 4). Further validation using IRT also indicated excellent fit of each sub-scale (Supplement 5).

The internal consistency reliability was excellent with a Cronbach's alpha of 0.94 for the HiSQOL total scale. Each of the three sub-scales also had excellent internal consistency reliability with Cronbach's alpha of 0.81-0.88 (Table 3). Test-retest reliability was also excellent for the HiSQOL total scale and each of the three sub-scales (Table 3). The hypotheses related to convergent validity were confirmed as the HiSQOL demonstrated very strong correlations between the HiSQOL total score and DLQI score (0.90). This is further supported by significant differences in HiSQOL mean score across disease severity bands for the DLQI (Figure 1). Additionally, the symptoms and psychosocial subscales had moderate convergent validity with the NRS for pain and HADS scores, respectively (Table 3).

Discussion

Development of an HS-specific HRQOL instrument has identified different aspects of HRQOL experienced by adults with HS, some of which are distinct from those captured by existing generic skin HRQOL tools such as the DLQI. For example, one of the major themes relates to drainage and odor, which are not found in the DLQI. The HiSQOL© is a 17-item HS-specific HRQOL instrument with a 7-day recall period. Expert HS clinicians and people with HS provided guidance and oversight throughout the process to ensure content validity. Items were generated from qualitative research with HS patients in two countries to ensure the most important constructs were included using patient-friendly language. Item selection took into account the qualitative findings, clinical importance, statistical analyses, and the need for the instrument to apply to a variety of participants in clinical trials regardless of age, sex, or location of HS disease activity. The final HiSQOL© instrument included items grouped into key sub-scales, organized around symptom, psychosocial, and functional concepts. It is important to note that the HiSQOL© total score and each sub-scale score relating to symptoms, psychosocial, and activities-adaptations can be used. Importantly, the test–retest reliability was strong and demonstrated stability of the HiSQOL© score when disease severity remained unchanged. Of the three instruments used to assess convergent criterion validity, the strongest correlation was between HiSQOL© and DLQI (r = 0.90), which is expected as they assess similar constructs and sample population (adults with skin disease). The psychometric assessment of the HiSQOL© also provided evidence on the discriminatory ability of the HiSQOL© by demonstrating significant differences in the HiSQOL© score across DLQI score bands[47].

The HiSQOL© differs from existing HS-specific HRQOL instruments[10]. It has 17-items separated into 3 sub-scales, that can be used independently or to generate a total score. The Hidradenitis

Suppurativa Burden Of Disease (HSBOD) is a 19-item instrument with responses on a 10-cm visual analog scale[53]. The HSBOD is divided into two parts with different recall periods: the last 4 weeks (14 items) and the entire time of having HS (5 items). The HSBOD internal consistency and convergent validity were compared against the DLQI with 29 HS patients, but full psychometric evaluation was not published. The instrument does not have validated sub-scales. The Hidradenitis Suppurativa Symptom Assessment (HSSA)-24 hour and HSSA-7 day are 9-item instruments with a 24-hour or 7-day recall period[54]. The HSSA instruments assess severity of symptoms and signs on an 11-point NRS scale and were preliminarily shown to be valid and reliable but a full psychometric evaluation was not published[54]. The Hidradenitis Suppurativa Impact Assessment (HSIA) is a 17-item instrument with a 7-day recall period and evaluates impacts of HS, but a full psychometric evaluation is also not published[54]. Sisic et al[55] developed an instrument called the Hidradenitis Suppurativa Quality of Life (HS-QoL) measure, which has a 6-month recall period and 44 items. Validation was assessed through pilot testing, but full psychometric analyses of the instrument structure were not performed[56]. Thorlacius et al[18] also performed preliminary work to develop an HS-specific HRQOL measure. Further development of these two instruments was curtailed to amalgamate efforts develop the HiSQOL.

Regarding study limitations, the participants in this study were drawn from referral practices and selection or response bias may limit generalizability. The majority of participants were Caucasian, with an underrepresentation of people with different races, ethnicities, or cultural beliefs that may influence responses to the instrument. However, efforts were made to recruit a broad sample of participants. The instrument demonstrates some floor effects and DIF. DIF analyses were only conducted for age and sex, so future studies will need to assess for DIF. There are several properties of the HiSQOL that remain to be elucidated including the responsiveness, minimal important difference and time to complete. While the HiSQOL was developed from patient interviews in two countries, further work is needed to confirm cross-cultural validity. Future studies are needed for adolescents, since HS can begin with or after puberty[57,58].. Although the HiSQOL© was developed for use in clinical trials, future studies will evaluate a reduced version of the HiSQOL© (HiSQOL-mini©). In summary, the HiSQOL© proved to be acceptable, comprehensible, and has strong evidence for validity and reliability in assessing patient-centered outcomes in clinical trials.

Table 1. Description of samples for concept elicitation, pilot testing, and psychometric assessment

	Concept Elicitation		Cognitive Debriefing 1		Cognitive Debriefing 2		Development sample
	US	**DK**	**US**	**DK**	**US**	**DK**	**US**
Total participants, n	21	21	15	15	15	15	222
Age, mean (range)	46.8 (23-74) years	37.9 (19-63) years	44.2 (21-73) years	37.0 (18-77) years	43.9 (25-68) years	42.3 (24-77) years	39.6 (range 19-77) years
Sex, n (%)							
Female	16 (76%)	13 (62 %)	13 (87%)	10 (67 %)	11 (73%)	11 (73%)	193 (87%)
Male	5 (24%)	8 (38 %)	2 (13%)	5 (33 %)	4 (27%)	4 (27%)	29 (13%)
Race							
White	14 (67%)	21 (100 %)	10 (66%)	NC	9 (60%)	NC	158 (712%)
Black	3 (14%)	--	5 (33%)		6 (40%)		48 (22%)
Asian	1 (5%)	--	--		--		3 (1%)
Bi- or Multiracial	3 (14%)	--	--		--		13 (6%)
Ethnicity		--					
Hispanic	3 (14%)	--	0 (0%)	NC	3 (20%)	NC	6 (3%)
Non-Hispanic	19 (86%)		15 (100%)		12 (80%)		216 (97%)
Hurley Stage*							
I, n (%)	0	3 (14%)	NC	NC	NC	NC	41 (19%)
II, n (%)	12 (57%)	12 (57%)					94 (42%)
III, n (%)	9 (43%)	6 (29%)					42 (19%)
							Not reported (45, 20%)

US: United States samples, DK: Denmark samples, NC: not collected

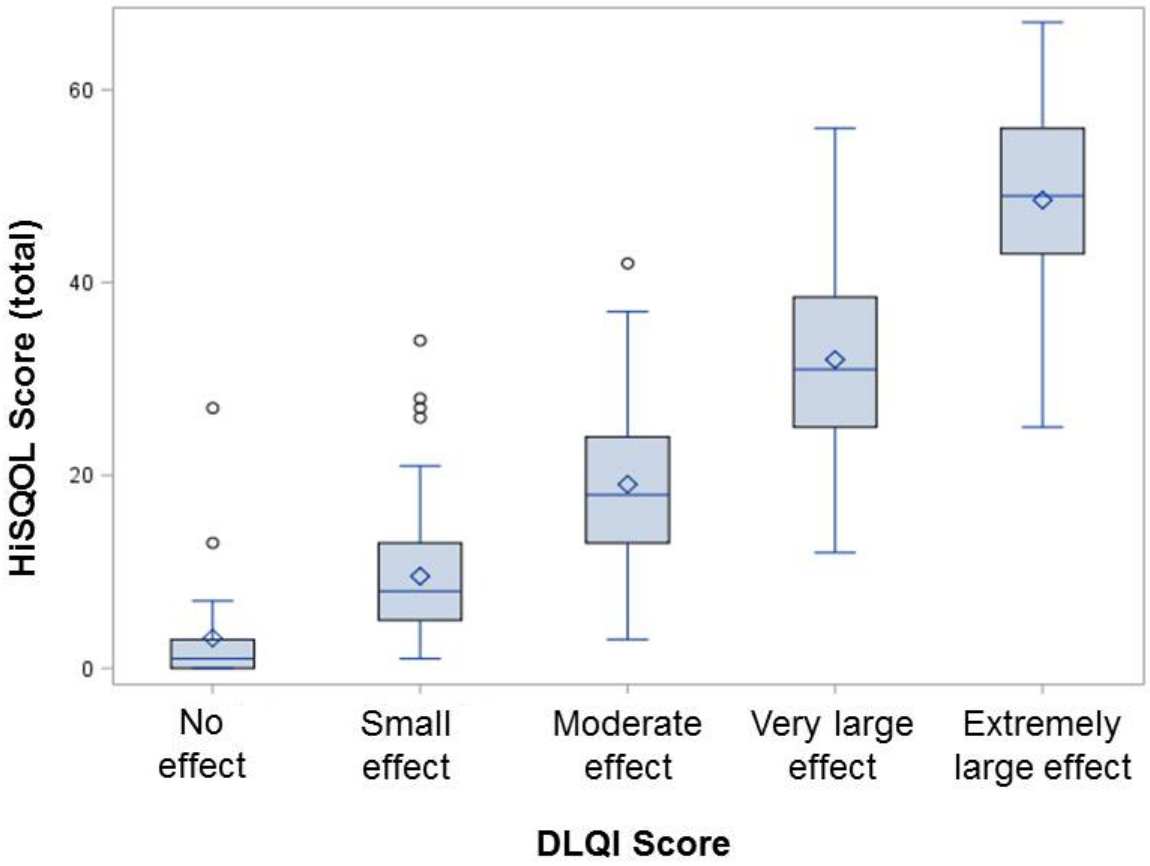
*Hurley stage is a disease severity staging score that categorizes the worst site of HS for a participant based on the presence of scarring, fistulas (or tunnels), and confluence of lesions[59].

Table 2. Results of descriptive statistics of items and confirmatory factor analyses of the US and DK samples.								
Development sample (US-based)								
Sub-scale	**Item**	**Mean**	**SD**	**Inter-item correlation**	**Floor**	**Ceiling**	**Item-total correlation**	**Factor Loading**
Symptoms	Pain	2.1	1.3	(0.55 to 0.65)	12.1	17.9	0.70	0.81
	Itch	0.9	1.2	(0.35 to 0.55)	54.3	4.7	0.49	0.55
	Drainage	1.9	1.4	(0.35 to 0.67)	19.6	18.8	0.69	0.90
	Odor	1.7	1.4	(0.36 to 0.67)	26.8	16.7	0.64	0.81
	CFA: Chi-Square=11.1, df=10, p=0.3464. RMSEA=0.023 (90% CI 0.000 to 0.078), CFI=0.999, TLI=0.998.							
Psychosocial	Down or depressed	1.7	1.4	(0.54 to 0.66)	25.3	14.0	0.75	0.90
	Embarrassed	2.3	1.5	(0.40 to 0.66)	18.0	30.1	0.71	0.85
	Anxious or nervous	1.5	1.4	(0.41 to 0.64)	34.5	13.6	0.68	0.80
	Concentration	1.0	1.1	(0.39 to 0.58)	40.7	2.7	0.58	0.65
	Sexual desire	2.2	1.6	(0.39 to 0.58)	24.6	35.2	0.58	0.71
	CFA: Chi-Square=26.9, df=15, P=0.0292. RMSEA=0.059 (90% CI 0.019 to 0.095), CFI=0.991, TLI=0.985.							
Activities-adaptations	Walking	1.2	1.1	(0.32 to 0.63)	35.7	3.1	0.68	0.73
	Exercising	1.8	1.5	(0.39 to 0.63)	27.1	19.0	0.72	0.79
	Sleeping	1.3	1.3	(0.28 to 0.62)	36.0	8.9	0.68	0.80
	Washing yourself	1.4	1.2	(0.38 to 0.68)	29.1	6.2	0.68	0.73
	Getting dressed	1.3	1.3	(0.41 to 0.68)	31.5	5.1	0.72	0.84
	What you wear	2.4	1.3	(0.34 to 0.61)	9.3	26.4	0.62	0.75
	Ability to work/study	1.3	1.4	(0.26 to 0.54)	44.2	14.3	0.60	0.68
	Sexual activity difficult	1.8	1.7	(0.36 to 0.54)	38.0	28.3	0.65	0.76
	Chi-Square=56.9, df=38, P=0.0286, RMSEA=0.046 (90% CI 0.015 to 0.070), CFI=0.992, TLI=0.989.							
Validation sample (DK-based)								
Sub-scales	**Item**	**Mean**	**SD**	**Inter-item correlation**	**Floor**	**Ceiling**	**Item-total correlation**	**Factor Loading**
Symptoms	Pain	2.0	1.3	(0.51 to 0.67)	12.3	14.9	0.72	0.80
	Itch	1.1	1.3	(0.35 to 0.51)	47.4	7.9	0.46	0.55
	Drainage	1.8	1.3	(0.35 to 0.68)	19.9	14.9	0.71	0.91
	Odor	1.6	1.4	(0.35 to 0.68)	28.0	13.6	0.65	0.82
	Chi-Square=8.4, df=8, p=0.3973, RMSEA=0.015 (90% CI 0.000 to 0.083), CFI=1.000, TLI=0.999							
Psychosocial	Down or depressed	1.6	1.4	(0.51 to 0.65)	28.3	12.1	0.75	0.89
	Embarrassed	1.9	1.5	(0.43 to 0.62)	25.3	22.3	0.71	0.73
	Anxious or nervous	1.2	1.4	(0.43 to 0.65)	43.9	9.8	0.69	0.90
	Concentration	1.0	1.1	(0.40 to 0.61)	42.2	4.0	0.61	0.87
	Sexual desire	2.1	1.6	(0.40 to 0.57)	25.9	32.8	0.58	0.60
	Chi-Square=27.5, df=15, p=0.0252. RMSEA=0.062 (90% CI 0.022 to 0.099), CFI=0.991, TLI=0.986.							
Activities-adaptations	Walking	1.1	1.2	(0.41 to 0.60)	40.3	4.3	0.65	0.77
	Exercising	1.8	1.5	(0.46 to 0.60)	30.2	20.2	0.74	0.85
	Sleeping	1.3	1.3	(0.43 to 0.61)	38.8	9.5	0.69	0.78
	Washing yourself	1.5	1.2	(0.40 to 0.73)	24.1	7.3	0.69	0.82
	Getting dressed	1.4	1.2	(0.40 to 0.73)	29.9	5.6	0.74	0.85
	What you wear	2.3	1.4	(0.39 to 0.54)	11.8	24.8	0.59	0.70

	Ability to work/study	1.1	1.4	(0.39 to 0.52)	52.5	12.5	0.61	0.75
	Sexual activity difficult	1.6	1.7	(0.41 to 0.48)	43.2	24.8	0.57	0.58
	Chi-Square=52.9, df=40, p=0.0829. RMSEA=0.039 (90% CI 0.000 to 0.065), CFI=0.995, TLI=0.994.							
SD: standard deviation								

Table 3. Reliability and convergent validity of the HiSQOL Sub-Scales and HiSQOL Total Scale				
	Symptom sub-scale	Psychosocial sub-scale	Activities-Adaptations sub-scale	HiSQOL total scale
Test-retest reliability (ICC) and Internal consistency (α)				
Cronbach's alpha (α)	0.81	0.85	0.88	0.94
Test-retest correlation (ICC)	0.85	0.84	0.90	0.90
Convergent validity				
DLQI			0.87	0.90
NRS for pain	0.74			
HADS Anxiety		0.69		
HADS Depression		0.63		

Figure 1. Known groups validity of HiSQOL across established DLQI score groups



		n (%)	HiSQOL		p-value*
			Mean	SD	
DLQI score bands	**No effect**	22 (5.4%)	3.1	6.2	<.0001
	Small effect	83 (20.5%)	9.6	6.5	
	Moderate effect	101 (24.9%)	19.1	7.6	
	Very large effect	116 (28.6%)	32.0	9.5	
	Extremely large effect	83 (20.5%)	48.6	8.4	

*p-value for comparison of means across groups
Based on complete case analysis, n=405 for combined US and DK responses

References

1. Buimer MG, Wobbes T, Klinkenbijl JH. Hidradenitis suppurativa. *The British journal of surgery.* 2009;96(4):350-360.
2. Garcia Martinez FJ, Menchen L. Pathogenesis: common pathways between hidradenitis suppurativa and Crohn disease. *Actas dermo-sifiliograficas.* 2016;107 Suppl 2:13-20.
3. Rayner CR. Pathogenesis, clinical features and management of hidradenitis suppurativa. *Annals of the Royal College of Surgeons of England.* 1997;79(4):309.
4. van der Zee HH, Laman JD, Boer J, Prens EP. Hidradenitis suppurativa: viewpoint on clinical phenotyping, pathogenesis and novel treatments. *Exp Dermatol.* 2012;21(10):735-739.
5. Jemec GB, Heidenheim M, Nielsen NH. Hidradenitis suppurativa--characteristics and consequences. *Clinical and experimental dermatology.* 1996;21(6):419-423.
6. Matusiak L, Bieniek A, Szepietowski JC. Psychophysical aspects of hidradenitis suppurativa. *Acta Derm Venereol.* 2010;90(3):264-268.
7. Wolkenstein P, Loundou A, Barrau K, Auquier P, Revuz J. Quality of life impairment in hidradenitis suppurativa: a study of 61 cases. *J Am Acad Dermatol.* 2007;56(4):621-623.
8. Matusiak L, Bieniek A, Szepietowski JC. Hidradenitis suppurativa markedly decreases quality of life and professional activity. *J Am Acad Dermatol.* 2010;62(4):706-708, 708 e701.
9. Kouris A, Platsidaki E, Christodoulou C, et al. Quality of Life and Psychosocial Implications in Patients with Hidradenitis Suppurativa. *Dermatology (Basel, Switzerland).* 2016;232(6):687-691.
10. Chernyshov PV, Zouboulis CC, Tomas-Aragones L, et al. Quality of life measurement in hidradenitis suppurativa: position statement of the European Academy of Dermatology and Venereology task forces on Quality of Life and Patient-Oriented Outcomes and Acne, Rosacea and Hidradenitis Suppurativa. *J Eur Acad Dermatol Venereol.* 2019.
11. Theut Riis P, Thorlacius L, Knudsen List E, Jemec GBE. A pilot study of unemployment in patients with hidradenitis suppurativa in Denmark. *Br J Dermatol.* 2017;176(4):1083-1085.
12. Deckers IE, Janse IC, van der Zee HH, et al. Hidradenitis suppurativa (HS) is associated with low socioeconomic status (SES): A cross-sectional reference study. *J Am Acad Dermatol.* 2016;75(4):755-759.e751.
13. Thorlacius L, Cohen AD, Gislason GH, Jemec GBE, Egeberg A. Increased Suicide Risk in Patients with Hidradenitis Suppurativa. *J Invest Dermatol.* 2018;138(1):52-57.
14. 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.
15. Thorlacius L, Ingram JR, Garg A, et al. Protocol for the development of a core domain set for hidradenitis suppurativa trial outcomes. *BMJ open.* 2017;7(2):e014733.
16. Thorlacius L, Garg A, Ingram JR, et al. Towards global consensus on core outcomes for Hidradenitis Suppurativa research: An update from the HISTORIC consensus meetings I and II. *Br J Dermatol.* 2017.
17. McKenna SP. Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science. *BMC medicine.* 2011;9:86.
18. Thorlacius L, Esmann S, Miller I, Vinding G, Jemec GBE. Development of HiSQOL: A Hidradenitis Suppurativa specific Quality of Life Instrument. *Skin Appendage Disorders* 2019([in press]).
19. Janse IC, Deckers IE, van der Maten AD, et al. Sexual health and quality of life are impaired in hidradenitis suppurativa: a multicentre cross-sectional study. *Br J Dermatol.* 2017;176(4):1042-1047.

20. Deckers IE, Kimball AB. The Handicap of Hidradenitis Suppurativa. *Dermatologic clinics.* 2016;34(1):17-22.
21. Waller N, John MT, Feuerstahler L, et al. A 7-day recall period for a clinical application of the oral health impact profile questionnaire. *Clinical oral investigations.* 2016;20(1):91-99.
22. Schneider S, Choi SW, Junghaenel DU, Schwartz JE, Stone AA. Psychometric characteristics of daily diaries for the Patient-Reported Outcomes Measurement Information System (PROMIS(R)): a preliminary investigation. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2013;22(7):1859-1869.
23. Sadosky A, Dukes E, Evans C. Reliability of a 1-week recall period for the Medical Outcomes Study Sleep Scale (MOS-SS) in patients with fibromyalgia. *Health Qual Life Outcomes.* 2009;7:12.
24. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes.* 2006;4:79.
25. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology.* 2010;63(7):737-745.
26. Elkjaer M, Dinesen L, Benazzato L, Rodriguez J, Logager V, Munkholm P. Efficacy of Infliximab treatment in patients with severe Fistulizing Hidradenitis Suppurativa. *Journal of Crohn's & colitis.* 2008;2(3):241-245.
27. Marron SE, Gomez-Barrera M, Tomas-Aragones L, et al. Development and Preliminary Validation of the HSQoL-24 Tool to Assess Quality of Life in Patients With Hidradenitis Suppurativa. LID - S0001-7310(19)30092-4 [pii] LID - 10.1016/j.ad.2019.02.002 [doi]. (1578-2190 (Electronic)).
28. Food and Drug Administration (FDA). *Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.* Rockville, MD.2009.
29. Merriam SB. *Qualitative Research: A Guide to Design and Implementation.* San Francisco, CA: Jossey-Bass; 2009.
30. Ferrans CE, Zerwic JJ, Wilbur JE, Larson JL. Conceptual model of health-related quality of life. *Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing.* 2005;37(4):336-342.
31. Bakas T, McLennon SM, Carpenter JS, et al. Systematic review of health-related quality of life models. *Health Qual Life Outcomes.* 2012;10:134.
32. U. S. Department of Health and Human Services. HealthMeasures: PROMIS (Patient-reported outcome measurement information system). 2018; http://www.healthmeasures.net/explore-measurement-systems/promis. Accessed February 2016-September 2017.
33. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine.* 2000;25(24):3186-3191.
34. Van der Veer K, Ommundsen R, Hak T, Larsen KS. Meaning shift of items in different language versions. A cross-national validation study of the illegal aliens scale. *Quality and Quantity.* 2003;37:193-206.
35. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide.* Cambridge: Cambridge University Press; 2011.

36. Basra MK, Fenech R, Gatt RM, Salek MS, Finlay AY. The Dermatology Life Quality Index 1994-2007: a comprehensive review of validation data and clinical results. *Br J Dermatol.* 2008;159(5):997-1035.
37. Lewis V, Finlay AY. 10 years experience of the Dermatology Life Quality Index (DLQI). *The journal of investigative dermatology Symposium proceedings.* 2004;9(2):169-180.
38. Snaith RP. The Hospital Anxiety and Depression Scale. *Health Qual Life Outcomes.* 2003;1:29.
39. Snaith RP, Zigmond AS. The hospital anxiety and depression scale. *British medical journal (Clinical research ed).* 1986;292(6516):344.
40. Hjermstad MJ, Fayers PM, Haugen DF, et al. Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. *Journal of pain and symptom management.* 2011;41(6):1073-1093.
41. Harris P, Taylor R, Thielke R, Payne J, Gonzalez N, Conde J. Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;421(2):377-381.
42. Kreiner S, Christensen KB. Graphical Rasch Models. In: Mesbah M, Cole FC, Lee MT, eds. *Statistical Methods for Quality of Life Studies*. Boston: Spring Science+Business; 2002:187-203.
43. Kreiner S, Christensen KB. Validity and Objectivity in Health-Related Scales: Analysis by Graphical Loglinear Rasch Models. In: von Davier M, Carstensen CH, eds. *Multivariate and Mixture Distribution Rasch Models*. Boston: Springer; 2007:329-346.
44. Christensen KB, Makransky G, Horton M. Critical Values for Yen's Q3 : Identification of Local Dependence in the Rasch Model Using Residual Correlations. . *Applied Psychological Measurement.* 2017;41(3):178-194.
45. Akoglu H. User's guide to correlation coefficients. *Turkish journal of emergency medicine.* 2018;18(3):91-93.
46. Hays RD, Anderson R, Revicki D. Assessing reliability and validity of measurement in clinical trials. Quality of Life assessment in clinical trials: methods and practice. London: Oxford University Press 1998:169–182.
47. Hongbo Y, Thomas C, Harrison M, Salek M, Finlay A. Translating the science of quality of life into practice: What do dermatology life quality index scores mean? . *J Invest Dermatol.* 2005;125:659-664.
48. Comrey AL. Factor-analytic methods of scale development in personality and clinical psychology. *Journal of consulting and clinical psychology.* 1988;56(5):754-761.
49. Kline P. *A Handbook of Test Construction: Introduction to Psychometric Design.* United Kingdom: Routledge; 1986.
50. Nunnelly JC. *Psychometric theory.* 2nd ed. New York: McGraw Hill; 1978.
51. Altman DG, Bland J. Measurement in medicine: the analysis of method comparison studies. *Statistician.* 1983;32:307-317.
52. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *Jama.* 1995;273(1):59-65.
53. Pinard J, Vleugels RA, Joyce C, Merola JF, Patel M. Hidradenitis suppurativa burden of disease tool: Pilot testing of a disease-specific quality of life questionnaire. *J Am Acad Dermatol.* 2018;78(1):215-217.e212.

54. Kimball AB, Sundaram M, Banderas B, Foley C, Shields AL. Development and initial psychometric evaluation of patient-reported outcome questionnaires to evaluate the symptoms and impact of hidradenitis suppurativa. *The Journal of dermatological treatment.* 2018;29(2):152-164.
55. Sisic M, Kirby JS, Boyal S, Plant L, McLellan C, Tan J. Development of a Quality-of-Life Measure for Hidradenitis Suppurativa. *Journal of cutaneous medicine and surgery.* 2017;21(2):152-155.
56. McLellan C, Sisic M, Oon HH, Tan J. Preliminary Validation of the HS-QoL: A Quality-of-Life Measure for Hidradenitis Suppurativa. *Journal of cutaneous medicine and surgery.* 2018;22(2):142-146.
57. Garg A, Lavian J, Lin G, Strunk A, Alloo A. Incidence of hidradenitis suppurativa in the United States: A sex- and age-adjusted population analysis. *J Am Acad Dermatol.* 2017;77(1):118-122.
58. Vazquez BG, Alikhan A, Weaver AL, Wetter DA, Davis MD. Incidence of hidradenitis suppurativa and associated factors: a population-based study of Olmsted County, Minnesota. *J Invest Dermatol.* 2013;133(1):97-103.
59. Hurley H. Axillary hyperhidrosis, apocrine bromhidrosis, hidradenitis suppurativa and familial benign pemphigus. Surgical approach. . In: Roenigk R, Roenigk H, eds. *Dermatologic Surgery, Principles and Practice.* New York, New York: Marcel Dekker; 1989.