



Published in final edited form as:

Br J Dermatol. 2021 May ; 184(5): 905–912. doi:10.1111/bjd.19565.

Hidradenitis Suppurativa Area and Severity Index Revised (HASI-R): psychometric property assessment

N. Goldfarb^{1,2,*}, M.A. Lowes^{3,*}, M. Butt⁴, T. King⁴, A. Alavi^{5,†}, J.S. Kirby^{4,†}

¹Departments of Medicine and Dermatology, University of Minnesota, Minneapolis, MN, USA

²Departments of Medicine and Dermatology, Minneapolis Veteran Affairs Health Care System, Minneapolis, MN, USA

³The Rockefeller University, New York, NY, USA

⁴Department of Dermatology, Penn State Health Milton S. Hershey Medical Center, Hershey, PA

⁵Department of Medicine, Division of Dermatology, Women's College Hospital, University of Toronto, Toronto, Canada

Abstract

Background: Validated, reliable, globally-accepted outcome measurement instruments for hidradenitis suppurativa (HS) are needed. Current tools to measure the physical signs domain for HS rely on lesion counts that are time-consuming and unreliable.

Objectives: To assess the reliability and validity of the Hidradenitis Suppurativa Area and Severity Index Revised (HASI-R) tool, a novel method for assessing HS severity, incorporating signs of inflammation, and body surface area (BSA) involved.

Methods: The measurement properties of the HASI-R tool were evaluated. The tool was created by combining the previously published HASI and Severity and Area Score for Hidradenitis (SASH) instruments. Twenty raters evaluated 15 HS patients in a **hospital-based ambulatory dermatology clinic**. The objectives of the study were to assess inter-rater and intra-rater reliability of the HASI-R and its components, as well as its construct and known groups validity. Existing lesion count-based clinician-reported measures of HS and their components were also assessed. Raters were also asked their preferences regarding the various HS severity assessment tools.

Results: The HASI-R had moderate inter-rater reliability (ICC = 0.60). This was better than all other HS physical sign outcome measures evaluated, which had poor inter-rater reliability (ICC < 0.5). HASI-R had the highest intra-rater reliability (ICC = 0.91). The HASI-R had good construct

Corresponding Author: Noah Goldfarb, MD, gold0414@umn.edu.

[†]These authors contributed equally

[†]These authors contributed equally

Authorship Disclosure: JSK has been a speaker for AbbVie, advisory board participant for AbbVie, consultant to AbbVie, Incyte, UCB, and ChemoCentryx, and participated in clinical trials with AbbVie, ChemoCentryx, Incyte, InflaRx, Novartis, and UCB. AA has undertaken personal advisory work and participated in the clinical trial with Pfizer, AbbVie, Janssen, UCB, Novartis, Celgene, LEO, Lilly, Sanofi, Valeant and Galderma. MAL has participated in advisory boards for Abbvie and Janssen and consulted with Incyte, Xbiotech, BSN, Kymera and Almirall. NG has participated in clinical trials with AbbVie, Chemocentryx and Pfizer. MB, TK have no conflicts of interest.

validity and demonstrated known groups validity. The HASI-R was also the most preferred tool by all raters.

Conclusions: Results from the clinometric assessment of the HASI-R are encouraging, and support continued evaluation of this clinician-reported outcome measure.

INTRODUCTION

Disease outcome measurement instruments that are valid, reliable, feasible and sensitive to change are critical for evaluating severity of a disease, response of a disease to therapy and comparing among existing and novel treatment options. Unfortunately, clinical studies in hidradenitis suppurativa (HS) have used varying outcome measurements, making comparisons between treatments and meta-analyses of current studies challenging. In a 2016 Cochrane review, 30 outcome measures were identified from twelve randomized controlled trials that were included in the analysis.¹ Importantly, ninety percent of these outcome measures lacked any validation data. Since then, there have been several validated tools developed to measure the physical signs of HS including the validated disease response endpoint Hidradenitis Suppurativa Clinical Response (HiSCR)² and International Hidradenitis Suppurativa Severity Score System (IHS4)³. Validation of older tools, such as the modified Sartorius scores and Hurley staging, have also been completed.⁴

HiSCR, developed for the PIONEER I and II studies, was validated as part of the study and is now the most popular tool for clinical outcome measurements for HS clinical trials.⁵ It was created as an end-point to evaluate response to treatment, but as it was created to only measure change over time, it is unable to measure cross-sectional disease severity. The IHS4, on the other hand, was created as a dynamic tool to measure severity as well as response to treatment.³ The components of the IHS4 include number of inflammatory nodules, abscesses and draining tunnels, each with predetermined weights. Hurley staging is a commonly used classification system that is relatively time-efficient and easy to use; however, it is not sensitive to changes in severity.⁶ Finally, the modified and original Sartorius scores also involve counting various lesion types with predetermined weights, measuring the longest distance between two relevant lesions within each anatomical region, and determining whether each region is Hurley stage III.^{7,8}

The first article to evaluate the reliability of multiple HS outcome measures found disappointing results.⁴ The intra-class correlation coefficients (ICC) for lesion counts ranged from poor for abscesses (ICC = 0.07), to fair for inflammatory nodules (ICC = 0.40), to moderate for draining tunnels (ICC = 0.46).⁴ The inter-rater reliability of the tools based on lesion counts was also problematic, with the baseline HiSCR (also known as the abscess nodule [AN] count) (ICC = 0.44) and IHS4 (ICC = 0.47) having only fair inter-rater reliability.⁴ More recent studies demonstrated moderate (ICC = 0.69)⁹ and high (ICC > 0.75)¹⁰ inter-rater reliability for the IHS4. The latter study showed that training improves inter-rater reliability. The addition of ultrasound has also been shown to improve inter-rater reliability, but it is not currently widely accessible amongst dermatologists.¹¹

In 2018, a consensus document was developed for outcome assessments in HS clinical trials and recommended assessment of lesion counts, involved anatomic locations, and surface

area.¹² This raises two issues for HS outcome measures: (1) the vast majority of the outcome instruments rely on lesions counts and do not assess body surface area (BSA), and (2) lesion counts have been shown to be problematic. Lesion counts are time-consuming and it can be difficult to define and discriminate among types of lesions based on clinical exam.¹³

In an attempt to move away from the challenges of counting lesions and include an assessment of anatomic location and BSA, the original Hidradenitis Suppurativa Activity and Severity Index (HASI) and Severity and Area Score for Hidradenitis Suppurativa (SASH) were created. The original HASI tool was a novel tool, based on the concept of the Psoriasis Area and Severity Index (PASI), that incorporated the signs of HS inflammation (erythema, thickness, drainage and tenderness) of various body locations with the estimated involved BSA.¹⁴ The SASH was a similar tool created by a separate group, that included inflammatory color change, induration, and amount of open skin surface with an estimate of involved BSA.¹⁵ The SASH was developed from qualitative interviews and focus groups of clinicians.¹⁵ Both tools' preliminary data suggested good reliability with strong correlation with existing measures, supporting their convergent construct validity. However, both tools had limitations including lack of tunnel assessment and incorporation of constructs like drainage and pain that are better reported by people with HS (i.e. patient-reported outcomes) rather than clinicians.

Since these tools were created simultaneously with the same goal, these groups merged to create a novel tool that addressed issues with the prior instruments, under the umbrella of the international Hidradenitis Suppurativa cORE outcomes set International Collaboration (HISTORIC) initiative.¹² The HASI revised (HASI-R) measures inflammatory color change, inflammatory induration, open skin surface, and extent of tunnels, in various body sites using an estimation of involved BSA. The objective of this study was to evaluate the inter-rater reliability, intra-rater reliability, convergent and known groups validity of the updated HASI-R tool.

PATIENTS AND METHODS

This study was approved by the Penn State Health Milton S Hershey Medical Center ethics board (IRB # STUDY00006806) as well as the Central IRB at Johns Hopkins (IRB # IRB00118850). All participants provided informed written consent.

HASI Revision

The HASI revision was guided by data collected by the two groups who created the original HASI and SASH. This included semi-structured HS patient interviews, a review of the literature on severity assessment tools, the consensus on core outcome set domains for HS clinical trials¹⁴ and suggestions from the US Food and Drug Administration (FDA). Discussion with regulatory authorities suggested a measure of extent of tunnels should be included as tunnels are a unique and core component of the HS disease process. Two main components are incorporated into the HASI-R, site-specific scores of disease severity and a BSA ordinal score for each involved anatomical region.

The HASI-R tool (Figure S1) includes four domains to assess the severity of HS disease activity including, inflammatory color change, inflammatory induration, open skin surface and extent of tunnels. Each of these variables is scored on a Likert scale from 0 to 3 (0 = none; 1 = limited/mild, 2 = moderate, 3 = severe/extensive) based on the average intensity for each body site. The term inflammatory color change refers to the pink to red spectrum of erythema in HS lesions, as well as the violaceous to purple spectrum, especially prominent in skin of color. The inflammatory induration domain refers to the inflammatory component of skin thickness. Open skin surface was defined as erosions, ulcerations or protuberant granulation-like tissue. As counting tunnels has been shown to have poor inter-rater reliability, the instrument was designed to measure the extent of tunnels on a Likert scale of 0–3. The maximum intensity score for each body site is 12 (maximum score of 3, across 4 components).

The HASI-R assesses HS activity at ten body sites including: head/neck, left axilla, right axilla, chest, abdomen, back, buttocks including intergluteal cleft, right thigh, left thigh, and pubic area/genitals. Extent of HS is assessed using BSA for each body site. Raters were asked to only score active disease, identified using the clinical signs of inflammation, including erythema and induration, as well as patient-reported symptoms of pain or drainage. Then, raters estimated BSA using the entire hand area (palm plus fingers), as the closest representation of 1% BSA in adults.^{16,17} Raters could provide BSA estimates to a tenth of a percentage point (e.g. 1.2%), using the assumption that the forefinger is approximately 0.1% BSA. The BSA score for each HS-specific site is calculated as the proportion of the site involved by HS. For example, if the left axilla had 1.5% BSA involved and the maximum allowed is 2% BSA, then the calculation is: $1.5\% / 2.0\% = 0.75$. Next the proportion is multiplied by 100% (e.g. $0.75 \times 100\% = 75\%$) and converted to a 0–6 ordinal scale, based on the psoriasis lattice system assessment (0 points = 0 BSA, 1 point = 1–3%, 2 points = 4–9%, 3 points = 10–20%, 4 points = 21–29%, 5 points = 30–50%, 6 points = >51%).¹⁸ These calculations were not performed by raters at the bedside, rather calculations were performed after the session, inputting raters' scores from paper evaluation forms.

To calculate HASI-R, the site-specific score is calculated by multiplying the BSA ordinal score by the sum of the four intensity scores. The maximum score for each body site is 72. Similar to the original SASH, no weighting was done for various sized regions.¹⁵ HASI-R score is the total for all body sites and can range from 0 to 720, with higher scores indicating more severe disease activity.

Assessment of the Measurement Properties

The HASI-R was evaluated in a multi-rater study to investigate inter-rater and intra-rater reliability and construct validity (convergent/divergent). In addition, known groups validity was also examined, in which the HASI-R scores were compared with mild, moderate and severe IHS4 scores. Participants with a broad range of HS severity and affected body sites were recruited and consented. For this early evaluation, only those patients with Fitzpatrick skin types I-IV were included. This study was conducted on a single day in March 2019. All clinicians completed a 90 minute training session, which included a section on the background and purpose of the study, a slide on HS terminology (e.g. nodule, abscess,

tunnel)^{14,19}, and slides on each of the HS severity tools used in the study, including the HASI-R, Hurley staging, and modified and original Sartorius scores. Ten picture sample cases were then reviewed for practice with the group, and the modified Sartorius, Hurley staging, and HASI-R were calculated for each case. Convergent construct validity for the HASI-R was assessed based on correlation with Hurley staging, modified and original Sartorius, AN count and IHS4. Every rater was randomized to score a patient a second time at the end of the session, to evaluate intra-rater reliability. Patients completed demographic questions and the Dermatology Life Quality Index (DLQI).²⁰ Divergent construct validity was assessed based on comparison of the HASI-R with the reverse-scored DLQI. Known group validity was determined across IHS4 severity groups.

After the session, raters were asked to rank the clinician-reported HS measures based on: (1) perceived ability to interpret the tool components/items (2) perceived ability to capture lesion severity; (3) perceived ability to capture HS extent; (4) perceived ability to capture meaningful change in HS during a clinical trial; and (5) preferred measure to use in a clinical trial.

Statistical Methods

The numbers of raters and HS participants were chosen to have an 80% power to detect the primary outcome of an ICC for inter-rater reliability based on a null hypothesis of an ICC of 0.5 and an alternative hypothesis of an ICC of 0.8 using an F-test with a significance level of 0.05. Descriptive statistics, including counts, means and standard deviations (SD), were calculated for the demographic data and survey responses. Inter-rater reliability was determined for each of the classification instruments and physical signs outcome measures as well as the individual components of the HASI and lesion types. An ICC > 0.9 was considered excellent, 0.76–0.89 high, 0.5–0.75 moderate, and less than 0.5 was considered poor reliability.²¹ The percent agreement for categorical ratings was calculated as the percentage of patients for which all the raters agreed. The agreement for continuous measures was calculated by taking the mean score for each patient (x-axis) and the differences from the mean score for each patient (y-axis) and calculating the 95% limits of agreement, defined as $\pm 1.96 \times SD$ around the mean score. To evaluate construct validity, partial Spearman's correlation coefficients were calculated between HASI-R and other outcome measures, adjusting for multiple provider scores. A correlation coefficient > 0.8 or < -0.8 was considered strongly associated, between +/-0.5 and +/-0.79 was considered moderately associated, between +/-0.20 and +/-0.49 was considered weakly associated, and between 0.00 and +/- 0.19, was considered not associated.²² Known-groups validity was based on an analysis of covariance (ANCOVA) for IHS4 known groups with the HASI-R score as the dependent variable. Data were collected on paper forms and entered into REDCap (Research Electronic Data Capture, a secure, web-based application designed to support data capture for research studies) by two study team members and double checked for accuracy. Statistical analysis was performed using SAS statistical software (SAS Institute Inc., Cary, NC, U.S.A.). Missing data was assumed to be random, and no data imputation was performed.

RESULTS

Demographics

The 20 raters consisted of primarily dermatologists (90%), from four countries (United States, Canada, Denmark, and Wales), with experience ranging from less than one year in practice (35%) to more than 20 years (25%) (Table 1). One rater was an advanced practice clinician and one was a general surgeon. Both had extensive experience in caring for patients with HS. HS participants were primarily white (80%), female (80%) of non-Hispanic/latino(a) ethnicity (80%). Our HS participants had a broad range in severity, including all Hurley stages (Table 1).

Reliability

The HASI-R had moderate inter-rater reliability (ICC = 0.6), which was the highest value among the outcome measures, aside from the Hurley classification at the axilla (Table 2). Other outcome measures including the original and modified Sartorius score, IHS4, and AN count had poor inter-rater reliability, ranging from an ICC of 0.20 to 0.43. The HASI-R had excellent intra-rater reliability (ICC=0.91), the highest of all outcome measures evaluated (Table 2). Hurley staging, modified and original Sartorius scores, and IHS4 scores had high intra-rater reliability, while the AN count had only moderate intra-rater reliability.

The observed percentage of agreement for the classification/staging systems was very low, ranging from 0–20% (Table 2). The limits of agreement ranged from ± 18.76 for the AN count to ± 137.0 for the modified Sartorius, indicating the impact of random variation on scores. If there was no random variation, then the differences between the raters' observations would be near zero. Similarly, the minimal detectable change (MDC), a statistic that takes into account the interrater reliability and describes the change in score that is not due to measurement error, varied from 27.25 for the AN count to 200.71 for the modified Sartorius.

Convergent, divergent and known groups validity

The convergent validity of the HASI-R with other clinician-reported measures was good with correlations ranging from 0.51 to 0.81 (Table 3). Divergent validity, assessed by correlation with the reverse-scored DLQI, showed a weak, non-significant correlation. Known groups validity demonstrated a statistically significant difference between the mean HASI-R scores across each of the IHS4 severity groups (mild, moderate and severe) (Figure 1).

Rater feedback

Raters ranked the HASI as the highest for ease of interpreting components, ability to capture HS extent and lesion severity, as well as ability to capture meaningful change. The HASI was the preferred tool for clinical trials by all raters. Hurley staging was the second highest ranked for ease of interpreting components, though all raters felt it was the poorest for ability to capture meaningful change (Table 4).

DISCUSSION

The HASI-R was developed as a novel method of measuring HS severity to avoid issues related to lesions counts, including the time-consuming nature and poor inter-rater reliability. This instrument measures BSA involved at various anatomic locations, important variables currently missing from HS clinical trials. Using data from the original reports of the original HASI and SASH, these two tools were merged into the current HASI-R tool.

The inter-rater reliability for the HASI-R in the current study was moderate, but higher than lesion count-based measures. All outcomes measures evaluated in this study had wide limits of agreement, which was also seen in the recent study by Thorlacius et al.¹⁴ This is also reflected in the large MDC values. This suggests that considerable score changes are needed to assure real change rather than measurement error. Using outcome measures in clinical trials with lower interrater reliability, increases the risk for erroneous results and requires larger sample sizes to appropriately power studies.

The HASI-R was the only tool that demonstrated excellent intra-rater reliability. All the other tools had high intra-rater reliability, except for the AN count, which only had moderate intra-rater reliability. Since all the evaluations took place on a single day, the intra-rater reliability could have been over-estimated for all clinician-reported outcome measures evaluated. The HASI-R also demonstrated convergent construct validity based on the strong, statistically significant correlation with IHS4. The weak and statistically non-significant, correlation of the reverse-scored DLQI with the HASI-R is consistent with other HS clinician-reported outcomes in the literature, ranging from correlation coefficients of 0.27 with HiSCR to 0.48 with the modified Sartorius.¹ While, the HASI-R demonstrated known groups validity based on the significant differences in mean HASI-R scores among the established IHS4 severity groups (mild, moderate, severe), the sample size in the current study was too small to determine cut-off values for HASI severity groups.

The major limitations of this study were the lack of feasibility data, inability to assess HASI-R's responsiveness over time, as well as limited involvement of various Fitzpatrick skin types. Assessment of feasibility in clinical trials and reproducibility in broader groups of raters and people with HS, including skin of color, is needed. The responsiveness of the HASI-R will be established through prospective trials. In this setting, the minimal clinical difference and cut-off values for severity groups will be established.

Overall, this study demonstrates the challenges of clinically evaluating HS by counting lesions, which have also been highlighted by others.^{4,9,14,15} The HASI-R's relatively higher inter-rater reliability, intra-rater reliability, construct validity, and rater preference suggest the HASI-R is a valid, reliable outcome measure for severity assessment in HS that may add value to count-based measures. Additionally, the core outcome set for HS trials calls for an assessment of BSA, in addition to lesion counts. Thus, the HASI-R provides a validated tool to assess this aspect. High-quality outcome measures for HS are needed urgently, as lower reliability, higher limits of agreement, and high MDCs of instruments, increase the likelihood that a treatment effect will be obscured (false negative).²³ This could result in

discarding promising therapies due to failure to detect actual differences in the disease after treatment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We would like to thank all of the patients who contributed to this study as well as our clinician raters including Haley B Naik, Gregor BE Jemec, John R Ingram, John W Frew, Ronda S Farah, Lori Fiessinger, Megan H Noe, Robert G Micheletti, Hadar A Lev-Tov, Angel S Byrd, Lauren A Orenstein, Isabelle Delorme, Nicole Boyer, Iltefat Hamzavi, Stephanie R Goldberg, and Marc Bourcier.

Funding Sources: Dr Kirby received funding from the Agency for Healthcare Research and Quality for this research (K08HS024585) and the International Dermatology Outcome Measures (IDEOM) organization.

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

REFERENCES

- Ingram JR, Hadjieconomou S, Piguat V. Development of core outcome sets in hidradenitis suppurativa: systematic review of outcome measure instruments to inform the process. *Br J Dermatol.* 2016;175(2):263–272. [PubMed: 26873867]
- Kimball AB, Jemec GB, Yang M, et al. Assessing the validity, responsiveness and meaningfulness of the Hidradenitis Suppurativa Clinical Response (HiSCR) as the clinical endpoint for hidradenitis suppurativa treatment. *Br J Dermatol.* 2014;171(6):1434–1442. [PubMed: 25040429]
- Zouboulis CC, Tzellos T, Kyrgidis A, et al. Development and validation of the International Hidradenitis Suppurativa Severity Score System (IHS4), a novel dynamic scoring system to assess HS severity. *Br J Dermatol.* 2017;177(5):1401–1409. [PubMed: 28636793]
- Thorlacius L, Garg A, Riis PT, et al. Interrater agreement and reliability of outcome measurement instruments and staging systems used in hidradenitis suppurativa. *Br J Dermatol.* 2019;181(3):482–491.
- Kimball AB, Okun MM, Williams DA, et al. Two Phase 3 Trials of Adalimumab for Hidradenitis Suppurativa. *The New England journal of medicine.* 2016;375(5):422–434. [PubMed: 27518661]
- Zouboulis CC, Desai N, Emtestam L, et al. European S1 guideline for the treatment of hidradenitis suppurativa/acne inversa. *J Eur Acad Dermatol Venereol.* 2015;29(4):619–644. [PubMed: 25640693]
- Sartorius K, Emtestam L, Jemec GB, Lapins J. Objective scoring of hidradenitis suppurativa reflecting the role of tobacco smoking and obesity. *Br J Dermatol.* 2009;161(4):831–839. [PubMed: 19438453]
- Sartorius K, Lapins J, Emtestam L, Jemec GB. Suggestions for uniform outcome variables when reporting treatment effects in hidradenitis suppurativa. *Br J Dermatol.* 2003;149(1):211–213. [PubMed: 12890229]
- Zouboulis CC, Matusiak L, Jemec GBE, et al. Interrater and intrarater agreement and reliability in clinical staging of hidradenitis suppurativa/acne inversa. *Br J Dermatol.* 2019;181(4):852–854. [PubMed: 30968946]
- Włodarek K, Stefaniak A, Matusiak Ł, Szepietowski JC. Could Residents Adequately Assess the Severity of Hidradenitis Suppurativa? Interrater and Intrarater Reliability Assessment of Major Scoring Systems. *Dermatology (Basel, Switzerland).* 2020;236(1):8–14.
- Martorell A, Alfageme Roldán F, Vilarrasa Rull E, et al. Ultrasound as a diagnostic and management tool in hidradenitis suppurativa patients: a multicentre study. *J Eur Acad Dermatol Venereol.* 2019;33(11):2137–2142. [PubMed: 31124183]

12. 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(3):642–650. [PubMed: 29654696]
13. Lipsker D, Severac F, Freysz M, et al. The ABC of Hidradenitis Suppurativa: A Validated Glossary on how to Name Lesions. *Dermatology (Basel, Switzerland)*. 2016;232(2):137–142.
14. Goldfarb N, Ingram JR, Jemec GBE, et al. Hidradenitis Suppurativa Area and Severity Index (HASI): a pilot study to develop a novel instrument to measure the physical signs of hidradenitis suppurativa. *Br J Dermatol*. 2020;182(1):240–242. [PubMed: 31286486]
15. Kirby JS, Butt M, King T. Severity and area score for hidradenitis (SASH): a novel outcome measurement for hidradenitis suppurativa. *Br J Dermatol*. 2020;182(4):940–948. [PubMed: 31233623]
16. Rhodes J, Clay C, Phillips M. The surface area of the hand and the palm for estimating percentage of total body surface area: results of a meta-analysis. *Br J Dermatol*. 2013;169(1):76–84. [PubMed: 23448271]
17. Agarwal P, Sahu S. Determination of hand and palm area as a ratio of body surface area in Indian population. *Indian J Plast Surg*. 2010;43(1):49–53. [PubMed: 20924450]
18. Langley RG, Ellis CN. Evaluating psoriasis with Psoriasis Area and Severity Index, Psoriasis Global Assessment, and Lattice System Physician's Global Assessment. *J Am Acad Dermatol*. 2004;51(4):563–569. [PubMed: 15389191]
19. Freysz M, Jemec GB, Lipsker D. A systematic review of terms used to describe hidradenitis suppurativa. *Br J Dermatol*. 2015;173(5):1298–1300. [PubMed: 26054001]
20. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)--a simple practical measure for routine clinical use. *Clin Exp Dermatol* 1994;19(3):210–216. [PubMed: 8033378]
21. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155–163. [PubMed: 27330520]
22. Zou KH, Tuncali K, Silverman SG. Correlation and simple linear regression. *Radiology*. 2003;227(3):617–622. [PubMed: 12773666]
23. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276–282. [PubMed: 23092060]

What's already known about this topic?

- Reliability of physician-reported outcome measures for hidradenitis suppurativa (HS), comprised of counting lesions, have come under scrutiny
- Current physician-reported outcome measures lack assessment of body surface area (BSA)
- The original Hidradenitis suppurativa Area and Severity Instrument (HASI) and Severity and Area Score for Hidradenitis (SASH) are two HS severity assessment instruments combining signs of inflammation with BSA, dispensing of traditional lesion counts

What does this study add?

- The two groups that created the HASI and SASH merged the two instruments into a single tool called the Hidradenitis suppurativa Area and Severity Instrument Revised (HASI-R)
- Clinometric assessment of the HASI-R was completed, including inter-rater reliability, intra-rater reliability, convergent/divergent validity and known groups validity
- Results from the clinometric assessment are encouraging and support continued validation of the HASI-R

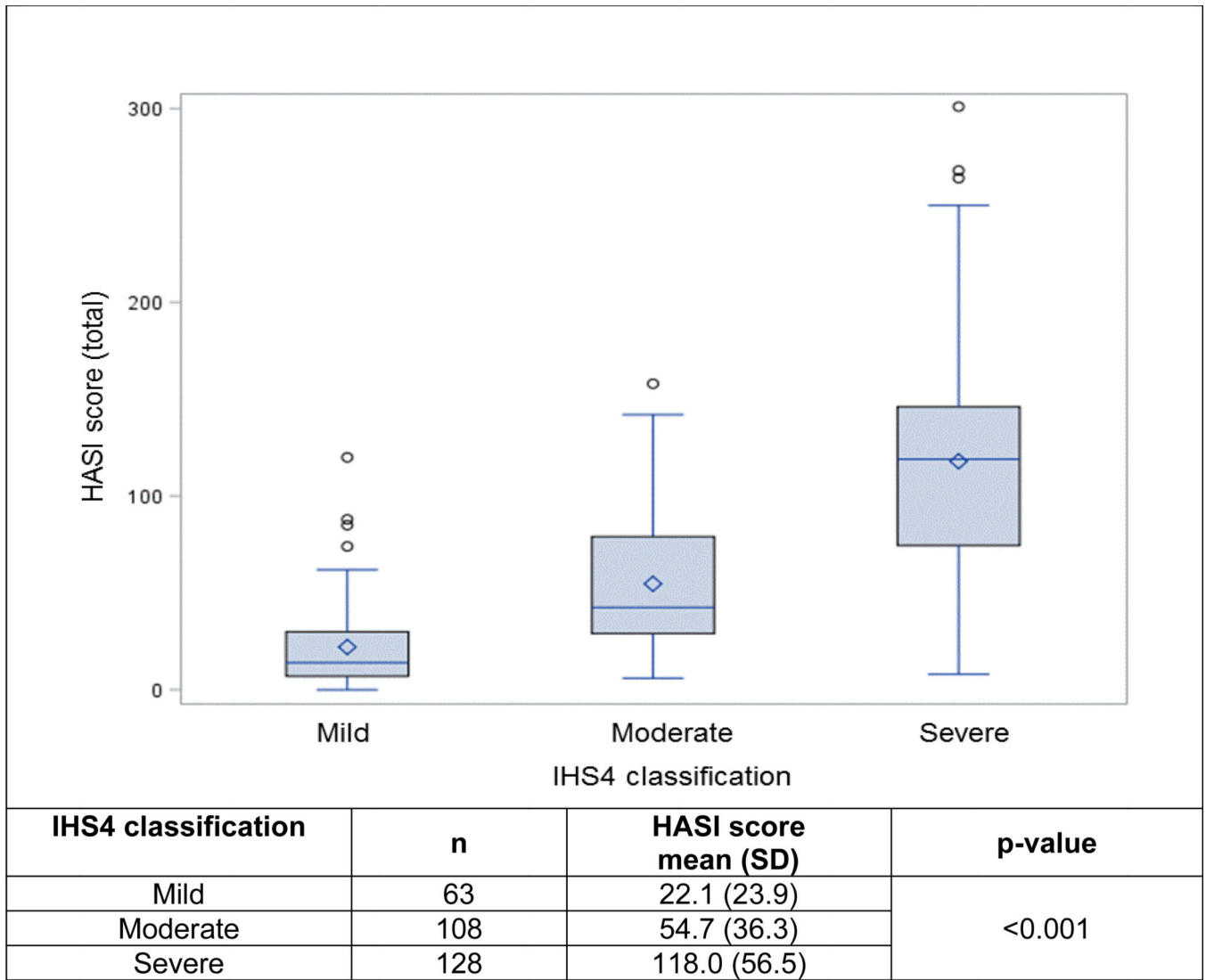


Figure 1.
Known groups validity of HASI-R across IHS4 score groups

Table 1.

Demographics of Raters and Participants with HS

Raters	
Sex, n (%)	
Male	8 (40%)
Female	12 (60%)
Provider Type, n (%)	
Dermatology MD	18 (90%)
Surgery MD	1 (5%)
APP	1 (5%)
Country of Practice, n (%)	
United States	15 (75%)
Canada	3 (15%)
Wales	1 (5%)
Denmark	1 (5%)
Years in Practice, n (%)	
Still in training	0
Less than 1 year	2 (10%)
1–5 years	7 (35%)
6–10 years	4 (20%)
11–20 years	2 (10%)
>20 years	5 (25%)
Participants with HS	
Sex, n	
Male	3 (20%)
Female	12 (80%)
Race, n (%)	
White	12 (80%)
Black/African-American	3 (20%)
Ethnicity, n (%)	
Not Hispanic or Latino(a)	12 (80%)
Hispanic or Latino(a)	1 (7%)
Missing	2 (13%)
Education, n	
Some High School	3 (20%)
High School Graduate	1 (7%)
Trade/ Technological/ Vocational	1 (7%)
Some College/ Associates	6 (40%)

Raters	
Undergraduate/Bachelor's	2 (13%)
Graduate	2 (13%)
Tobacco, n (%)	
Never Smoked	7 (47%)
Former Smoker	5 (33%)
Current Smoker	3 (20%)
BMI, mean (SD)	38.27 (8.11)

Abbreviations. AP, Advanced Practice Provider, BMI, body mass index; SD, standard deviation

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Reliability of the measures

Variable [possible range]	Rater Score Range In Study	Observed Agreement*	ICC (95%CI)	MDC	Missing Data, %
Inter-rater reliability					
Classification/Staging instruments					
Hurley** axillae [0–3]	0–3	2 (13%)	0.80 (0.66, 0.89)	1.35	6.2%
Hurley** groin [0–3]	0–3	0 (0%)	0.58 (0.39, 0.74)	1.86	5.2%
Hurley** gluteal [0–3]	0–3	3 (20%)	0.46 (0.28, 0.64)	1.90	4.3%
IHS4 classification [1–3]	1–3	3 (20%)	0.51 (0.32, 0.69)	1.47	2.0%
Physical signs outcome measures					
HASLER					
HASLER [0–648]	0–301	-71.33 to 71.33	0.60 (0.41, 0.76)	103.47	0.3%
IHS4 total [0-unlimited]	0–171	-42.15 to 42.15	0.43 (0.25, 0.62)	62.06	0.3%
Sartorius, modified [0-unlimited]	0–329	-109.60 to 109.60	0.35 (0.20, 0.55)	162.07	0.3%
Sartorius, original [0-unlimited]	0–463	-137.0 to 137.04	0.22 (0.10, 0.40)	200.71	0.3%
HISCR baseline (AN count) [0-unlimited]	0–93	-18.76 to 18.76	0.20 (0.09, 0.37)	27.25	0.3%
HASLER Components					
Extent of tunnels [0–3]			0.61 (0.42, 0.77)	0.71	0.3%
Inflammatory induration [0–3]			0.61 (0.42, 0.77)	0.74	0.3%
Inflammatory color change [0–3]	See Table S2	See Table S2	0.57 (0.38, 0.74)	0.73	0.3%
BSA ordinal score [0–6]			0.53 (0.34, 0.71)	1.25	0.3%
Open skin score [0–3]			0.50 (0.32, 0.69)	0.67	0.3%
Lesion Count Components					
Draining tunnel [0-unlimited]	0–27	-6.50 to 6.50	0.56 (0.37, 0.73)	9.59	0.3%
Total tunnel [0-unlimited]	0–45	-10.99 to 10.99	0.49 (0.30, 0.67)	16.32	0.3%
Inflammatory nodules [0-unlimited]	0–53	-13.91 to 13.91	0.21 (0.10, 0.38)	20.14	0.3%
Total nodules [0-unlimited]	0–89	-20.99 to 20.99	0.16 (0.07, 0.33)	30.52	0.3%

Variable [possible range]	Rater Score Range In Study	Observed Agreement*	ICC (95%CI)	MDC	Missing Data, %
Abscesses [0-unlimited]	0-40	-8.80 to 8.80	0.05 (0.01, 0.19)	12.87	0.3%
Intrarater reliability					
Measures					
HASI-R			0.91 (0.81, 0.96)		
Sartorius, modified			0.85 (0.69, 0.94)		
Sartorius, original			0.89 (0.75, 0.95)		
IHS4			0.86 (0.71, 0.94)		
Hurley (all sites)			0.80 (0.59, 0.91)		
HISCR baseline (AN count)			0.74 (0.50, 0.89)		

* Observed agreement is reported as percentage of sites for which all raters agreed, for continuous outcomes the Limits of Agreement is reported.

** The Hurley staging system does not include any clear stage in the original version, but in this study '0' was added to represent skin that did not have active HS lesions. Abbreviations: AN, abscess/nodule; BSA, body surface area; HASI, Hidradenitis suppurativa Activity Index; ICC, Intraclass Correlation Coefficient; IHS4, International HS Severity Scoring System; MDC, minimal detectable change

Table 3.

Convergent/Divergent validity of HASI-R with other physical sign outcome measures

Measures	HASI correlation coefficient (95CI) *
Convergent Validity	
IHS4	0.81 (0.77–0.85)
Hurley stage	0.79 (0.74–0.83)
Sartorius, modified	0.73 (0.67–0.78)
Sartorius, original	0.69 (0.62–0.74)
HiSCR baseline (AN count)	0.51 (0.41–0.59)
Divergent Validity	
DLQI (reverse-scored)	–0.45 (–0.78–0.08)

* Partial Spearman correlation coefficient adjusts for different providers

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4.

Median (IQR) rank of measures from rater post-session surveys

Please rank the tools (1–5)* based on:	HASI-R	Hurley stage	Original Sartorius	Modified Sartorius	IHS4
Ease of interpreting components	1.00 [1,2]	1.00 [1,2]	5.00 [5,5]	4.00 [4,4]	3.00 [2,3]
Ability to capture HS extent	1.00 [1,1]	5.00 [4,5]	3.50 [2,5]	3.00 [3,4]	3.00 [2,3]
Ability to capture severity of HS lesions	1.00 [1,1]	5.00 [4,5]	3.00 [3,4]	3.00 [2,3]	3.00 [2,4]
Ability to capture meaningful change in HS	1.00 [1,1]	5.00 [5,5]	3.00 [2,4]	3.00 [2,3]	3.00 [2,4]
Your preferred tool	1.00 [1,1]	5.00 [4,5]	4.00 [4,5]	3.00 [3,4]	2.00 [2,2]

* A score of 1 is most favorable and 5 is least favorable

Abbreviations. AN, abscess/nodule; HASI-R, Hidradenitis suppurativa Activity Index, Revised; IHS4, International HS Severity Scoring System; IQR, Interquartile range.