New treatments and new assessment instruments for Hidradenitis suppurativa

Kelsey R. van Straalen\(^1,2\) | John R. Ingram\(^2,3\) | Matthias Augustin\(^2,4\) | Christos C. Zouboulis\(^2,5\)

\(^1\)Department of Dermatology, University of Michigan, Ann Arbor, Michigan, USA
\(^2\)European Hidradenitis Suppurativa Foundation e.V., Dessau, Germany
\(^3\)Department of Dermatology, Division of Infection & Immunity, Cardiff University, Cardiff, UK
\(^4\)Institute for Health Care Services Research in Dermatology and Nursing, University Hospital Medical Center Hamburg-Eppendorf, Hamburg, Germany
\(^5\)Departments of Dermatology, Venereology, Allergology and Immunology, Dessau Medical Center, Brandenburg Medical School Theodor Fontane and Faculty of Health Sciences Brandenburg, Dessau, Germany

Abstract
Research interest in Hidradenitis Suppurativa (HS) has grown exponentially over the past decades. Several groups have worked to develop novel scores that address the drawbacks of existing investigator-assessed and patient-reported outcome measures currently used in HS trials, clinical practice and research. In clinical trial settings, the drawbacks of the HiSCR have become apparent; mainly, it is lack of a dynamic measurement of draining tunnels. The newly developed (dichotomous) IHS4 and HASI-R are backed up by adequate validation data and are good contenders to become the new primary outcome measure in HS clinical trials. Patient-reported outcomes, as well as physician reported measures, are being developed by the Hidradenitis Suppurativa Core outcomes set International Collaboration (HISTORIC). For example, the Hidradenitis Suppurativa Quality of Life (HiSQOL) score is a validated measure of HS-specific quality of life and is already being used in many HS trials. Magnitude of pain measurement via a 0–10 numerical rating scale is well-established; however, consensus is still required to ensure consistent administration and interpretation of the instrument. A longitudinal measurement over multiple days rather than at one time point, such as for example the Pain Index could provide increased reliability and reduced recall bias. Ultimately, these newly developed scores and tools can be included in a standardized registry to be used in routine clinical practice.

Keywords
ERHS, HiSCR, HiSQOL, HISTORIC, IHS4, outcome measurement instruments, pain, pain index, patient-reported outcome measures, quality of life

1 | INTRODUCTION
Hidradenitis Suppurativa is a chronic, debilitating inflammatory skin disease which affects around 1% of the Western population.\(^1\) Research interest in this disease has grown exponentially over the past decades. With the increasing number of clinical trials in HS come new insights into the drawbacks of existing investigator assessed and patients reported outcomes. Moreover, the development of several international clinical registries highlights the need for the implementation of (different) outcome measures in routine clinical care. This article provides an overview of the latest developments in both investigator and patient-reported outcomes for HS as discussed during the 11th EHSF Conference in 2022.
2 | NEW INVESTIGATOR-ASSESSED OUTCOMES

Hidradenitis suppurativa (HS) presents with a wide range of inflammatory and non-inflammatory lesions including (inflammatory) nodules, abscesses, and draining tunnels as well as open pseudocomedones, scars and ulceration. These lesions, in particular nodules and abscesses, know a high natural variability within patients. Capturing this highly heterogeneous disease in a single investigator-assessed outcome measure has proven difficult. As a consequence, to date over 20 different physician-assessed outcome measures have been proposed. Not only do most lack adequate validation data, this jungle of investigator-assessed outcomes has severely hampered comparison of published studies. The majority of scoring systems are developed to measure response to anti-inflammatory therapy and are less suitable to assess surgical interventions.

2.1 | HiSCR

The Hidradenitis Suppurativa Clinical Response (HiSCR) was proposed in 2014 and developed retrospectively from a randomized controlled trial (RCT) that used other outcome measures in the trial itself. HiSCR has since been adopted as an FDA-supported primary endpoint in almost all RCTs subsequently. The HiSCR identifies responders as those who achieve at least a 50% reduction in abscess and nodule count (AN-count) without an increase in the number of abscesses or draining tunnels relative to baseline. As this score only dynamically measures abscesses and nodules, patients with a AN-count under three were excluded from the development to ensure that a reduction of one abscess or nodule did not result in achieving the endpoint. However, this has subsequently led to the exclusion of patients with an AN-count <3 but many draining tunnels from current clinical trials that use HiSCR as the primary endpoint.

During recent clinical trials, other drawbacks of the HiSCR were identified. In particular the SHINE study, a phase II RCT assessing the efficacy of IFX-1 (vilobelimab) in patients with moderate–severe HS compared with placebo was instrumental in bringing these drawbacks to light. While participants in the highest dosed treatment group achieved a significantly greater reduction in AN-count and draining tunnels relative to the placebo group at Week 16, the HiSCR rate was not statistically different between these groups. This illustrates that the HiSCR, by not dynamically incorporating draining tunnels, does not fully capture the effect of anti-inflammatory treatment. Therefore, other outcomes have been developed that either dynamically take tunnels into account or provide a totally new perspective on scoring the inflammatory burden in HS; the IHS4, the dichotomous IHS4 and HASI-R, respectively.

2.2 | IHS4 and dichotomous IHS4

The International Hidradenitis Suppurativa Severity Score System (IHS4) is a continuous score that assigns different weights to different lesion types: inflammatory nodules (1 point), abscesses (2 points) and draining tunnels (4 points). Disease severity bands for IHS4 have been developed:≤ 3 points; mild, 4–10 points; moderate and ≥11 points; severe. The continuous IHS4 score has been adopted as a secondary outcome measure, in addition to the HiSCR, in many recently completed, ongoing and upcoming clinical trials. However, the preference of the FDA for a dichotomous outcome has resulted in the continuous IHS4 not being implemented as a primary outcome after the drawbacks of the HiSCR have surfaced. Therefore, as was presented during the 11th EHSF Conference in 2022, a dichotomous version of the IHS4 (IHS4-55) has been developed. The optimal cut-off threshold was identified as a 55% reduction in total IHS4 score. The performance of this IHS4-55 was presented to be similar to that of HiSCR in the PIONEER datasets while addressing the main drawbacks of the HiSCR; the dichotomous IHS4 takes draining tunnels into account in a dynamic and validated manner, and it does not exclude patients with an AN-count <3 but many draining tunnels. Moreover, the external validation of the dichotomous IHS4 in a large Europe wide prospective antibiotics study showed that the score was not only responsive in patients treated with adalimumab but also in patients treated with antibiotics.

Nonetheless, the IHS4 and its newly developed dichotomous version (and the HiSCR) have a well-recognized drawback; they rely on counting individual lesions. In patients with more severe disease, where lesions tend to coalesce, counts of individual lesions were shown to vary highly among raters compared with patients with milder disease. Potentially as a result of these difficulties, the AN-count of the HiSCR and the continuous IHS4 score only reach fair inter-rater reliability.

2.3 | HASI-R

To move away from counting lesions the Hidradenitis Suppurativa Area and Severity Index (HASI) and its revised version, the HASI-R were developed as part of the HISTORIC initiative. The HASI-R assesses 10 different body sites and scores the average intensity of skin discoloration due to inflammation, induration due to inflammation, extent of tunnel formation and the extent of open skin surface on a 4 point Likert scale: (0: clear, 1: mild/limited, 2: moderate and 3: severe/extentive) in each area. In addition, at each of the assessed body sites, the percentage of body surface area (BSA) involved by active HS is assessed and converted to a 0–6 ordinal scale per skin region (0 points for 0% BSA, 1 point for 1–3% BSA, 2 points for 4–9% BSA, 3 points for 10–20% BSA, 4 points for 21–29% BSA, 5 points for 30–50% BSA and 6 points for >51% affected BSA in that area). The total HASI-R score (ranging from 0 to 720) is calculated by...
multiplying the ordinal BSA score with the sum of the disease activity scores at each site and adding these together. With an intraclass correlation coefficient (0.60, 95% CI 0.41–0.76), the inter-rater reliability of the HASI-R was moderate. Raters identified the HASI-R as their preferred tool. However, it should be noted that the raters were not required to transform the BSA into the needed ordinal score or calculate the total score.

Nonetheless, this score might also have some drawbacks. Firstly, discoloration due to inflammation might be underestimated in patients with dark skin compared to those with lighter skin types. Secondly, especially in severe disease, cutaneous thickening is likely comprised of both induration due to active inflammation as well as scarring. Lastly, the score measures the extent of tunnel formation rather than drainage as a proxy for inflammation. For all intents and purposes, the baseline and surgical area measurements from the SHARPS study can serve as an illustration of how the extent of tunnel formation responds to anti-inflammatory therapy (adalimumab).

In this study, adalimumab did not significantly reduce the size of the area requiring surgery or in other words the extent of tunnel formation. Taken together, it remains unknown how the HASI-R performs in a diverse patient population and how responsive this score is to change after anti-inflammatory therapy. Studies to assess these qualities are underway.

3 | PATIENT-ASSESSED OUTCOMES

Patient-assessed outcomes for hidradenitis suppurativa (HS) are a key element of the Hidradenitis Suppurativa cORe outcomes set International Collaboration (HISTORIC). The mission of HISTORIC is to develop a core set of outcome measures for HS clinical trials by determining the core domains (the “what” to measure) and the measurement instrument for each domain (the “how” to measure). Inclusion of the core set in clinical trials going forward will prevent outcome measure heterogeneity and allow results from different trials to be compared in meta-analyses. Outcomes are not restricted to the core set and triallists may wish to include other instruments as well, depending on the intervention(s) under investigation.

The consensus process involves more than 100 patients, clinical experts and methodologists from 19 countries in 4 continents. To reach consensus regarding six core domains required five e-Delphi survey rounds and four face-to-face consensus meetings. Of the six core domains, four are patient-assessed outcomes, namely HS-specific quality of life, patient global assessment, pain and other symptoms (drainage and fatigue) (Figure 1). The symptom domain was strongly supported by patients and did not quite reach “consensus in” for the clinical experts; nevertheless, the domain is included in the core set because the HISTORIC Steering Committee felt that the patient voice should supersede clinicians for a patient-reported outcome.

Having determined the core set of domains, HISTORIC’s next challenge is to assess instruments currently available to measure each domain. In some domains, there are no available suitable instruments and so HISTORIC work groups are developing a new instrument. In other cases, pre-existing instruments are being assessed to determine whether they have sufficient validity, reliability, responsiveness and feasibility in the HS trial setting to be put forward as the instrument of choice.

3.1 | HiSQOL

No fully validated instruments existed to measure HS-specific quality of life and so a new measure was developed, using items highlighted from the core domains process. Content validity was sought from HS patient interviews in the United States and Denmark. The Hidradenitis Suppurativa Quality of Life (HiSQOL) score has 17 items grouped into three sub-domains covering activities, psychological impact and symptoms. It is scored using a Likert scale with responses ranging from no impact to extreme impact and with a recall window covering the previous 7 days. Importantly, being unable to undertake an activity is scored as having maximum impact on quality of life, rather than being not applicable. HiSQOL is embedded within
several industry and academic HS trials, the data from which will be used to provide additional validation evidence. The instrument is being considered by the FDA certification programme, and, to date, HiSQOL is translated into approximately 20 languages.

3.2 | Patient Global Assessment

The global assessment domain identified by HISTORIC has both Investigator Global Assessment and Patient Global Assessment (PtGA) sub-domains. The PtGA work group determined that there was not a sufficiently validated instrument already available and so a new instrument was needed. Interview studies determined that a global HS-specific quality of life single item has the required content validity. The final wording of the question is as follows: "In the past 7 days how much has your HS influenced your quality of life?" Answers are on a 5-point ordinal scale, comprising "Not at all, slightly, moderately, very much, or extremely." Good test–retest reliability has been demonstrated, as well as convergent validity with the Dermatology Life Quality Index and responsiveness to change.

3.3 | Pain

In the long list of more than 100 items initially identified by the HISTORIC domains process, pain was consistently ranked as the most important item. Hence, rather than being merged with other items, the pain item was adopted as a core domain by itself. On assessment of the pain literature, pain has three main facets: magnitude, functional effects and character. The pain work group decided that functional effects of pain are captured by HiSQOL and so should not be included in the pain instrument to avoid duplication. It was felt that pain character is harder to quantify in a clinical trial and measurement would be needed for only a subset of trials, rather than featuring in the core set. Taking forward the pain magnitude component, the most used instruments in painful skin conditions are the pain numerical rating scale (NRS) and visual analogue scale (VAS). The two instruments are equivalent and NRS was selected to measure drainage in HS. Several instruments already exist to measure fatigue in other conditions and these will be assessed to determine whether one can be used in HS, being careful to distinguish fatigue from sleep disturbance, which is a separate concept.

3.4 | Symptoms of drainage and fatigue

As well as pain, several other HS symptoms were identified during the core domains process, including drainage, odour, fatigue and pruritus. After careful prioritization during discussions at in-person HISTORIC consensus meetings, it was decided to include drainage and odour in the core set on the basis that odour is secondary to drainage and pruritus, while important, is not a universal symptom in HS. Work is currently underway to develop and validate an instrument to measure drainage in HS. Several instruments already exist to measure fatigue in other conditions and these will be assessed to determine whether one can be used in HS, being careful to distinguish fatigue from sleep disturbance, which is a separate concept.

4 | HOW TO MAKE PATIENT-ASSESSED OUTCOMES RELIABLE

The "patient centered" care movement in the 1970s led to interest in models of shared decision-making. Shared decision-making defines the goal of clinical communication as agreement between physicians and patients. However, available evidence shows that discussions around decisions in outpatient care lack essential components of information, namely correct recall of symptoms or misunderstanding of physician questions.

There is currently a large number of non-disease specific patient-reported outcome measure instruments applied in HS. These patient-reported outcome measures are commonly used in other dermatologic diseases, as such as Dermatology Life Quality Index (DLQI), Derriford Appearance Scale-24 and Work Productivity and Activity Impairment, has been reported to be less suitable for the assessment of HS, since they lack the sensitivity to reflect HS-associated changes in quality of life and frequently show a poor correlation with investigator-assessed outcome measure instruments and HS-specific patient-reported outcome measures (Table 1). This plethora of outcomes indicates the intensive, but until now futile efforts of developing an overall accepted patient-reported outcome. The poor correlation with investigator-assessed outcome measure instruments has been observed in many fields of medicine, for example the poor correlation of PASI and DLQI in psoriasis. Although this is not surprising, since they provide different constructs, this fact provides an obvious explanation for their ineffective inclusion in validated instruments, at least in HS. On the contrary, HS exhibits one of the highest reductions in quality of life among skin diseases, and it does not seem to improve at a similar magnitude as other aspects of the disease over time despite the emergence of effective treatments.

More than 75% of patients report physical symptoms, especially drainage and pain, but also skin irritation, itching, bleeding and odour, as the main factors reducing their quality of life. However, after 1 week, only 49% of patients recall the same information accurately without prompting, 36% with a prompt and 15% recall erroneously or not at all. Regarding the recall capacity, patient education also plays a role: Patients with less than high school education recall 38% of items accurately, while patients with a college degree recall a significantly higher rate (65%). Moreover, recall of skin damage-induced (nociceptive) pain is complicated by the presence of chronic central sensitization (neuropathic pain), which arises from dysregulation of the damage-reporting system, namely the central nervous system, in chronic diseases. The painDETECT questionnaire and the VAS and NRS pain in daily, weekly or monthly intervals have
TABLE 1 Patient-reported outcome measures applied in HS and HS-specific patient-reported outcome measures

<table>
<thead>
<tr>
<th>Patient-reported outcome measures applied in HS</th>
<th>HS-specific patient-reported outcome measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Dermatology Life Quality Index (DLQI)</td>
<td>• Quality of life for Hidradenitis Suppurativa (HS-QoL, 2017)</td>
</tr>
<tr>
<td>• Hidradenitis Suppurativa Symptom Assessment (HSSA)</td>
<td>• Hidradenitis Suppurativa Burden Of Disease (HSBOD, 2018)</td>
</tr>
<tr>
<td>• Patient Global Assessment of Hidradenitis Suppurativa Lesions</td>
<td>• Hidradenitis Suppurativa Impact Assessment (HSIA, 2018)</td>
</tr>
<tr>
<td>• Patient Global Assessment Visual Analogue Scale (VAS)</td>
<td>• HSQoL-24 (2019)</td>
</tr>
<tr>
<td>• Patient Global Assessment Scale</td>
<td>• HIDRAdisk (2019)</td>
</tr>
<tr>
<td>• Pain VAS</td>
<td>• Hidradenitis Suppurativa Quality of Life (HiSQLQ, 2020)</td>
</tr>
</tbody>
</table>

been included in clinical studies, all with the disadvantage of symptom severity recall by patients. Interestingly, the presence of chronic central sensitization worsens the capacity of acute pain recall after a period of time.23–25

One way to reduce these errors and make patient-reported pain a more reliable assessment would be by minimizing the importance of single assessments. Such an instrument, the Pain Index,36 has been developed for accurate prospective detection of nociceptive pain; it is based on the numerical rating scale (NRS; 0, no pain; 10, most severe pain) and represents the sum of daily NRS assessment over a period of 30 days (daily from Day 0 to Day 30). In a recent clinical study, the Pain Index has shown a high degree of correlation with NRS and a moderate correlation with IHS4, HS-PGA and DLQI.37 By including such instruments in for example a mobile phone app, daily prompting the patient to assess their clinical symptoms and/or the current daily pain scores at the same time, the accuracy of these longitudinal measurements would be increased.

5 | REAL WORLD OUTCOMES

The objective of clinical research is to investigate the efficacy and safety of a defined therapeutic measure, for example a drug therapy, beyond doubt and, if possible, under exclusion of confounding factors. As previously discussed, this requires precisely formulated endpoints that allow the research questions to be answered with the greatest possible validity and sensitivity.

In contrast, real-world outcomes in routine care under routine conditions have different objectives. They are used for decision-making in individual patients and thus to capture effects, side effects, and patient preferences as well as potential treatment barriers and economic factors. Unlike in clinical research, more individual factors go into these treatment decisions, such as patient preferences, comorbidity and comedication. In routine care, it is also necessary to examine the quality of care in terms of quality management with extended outcomes. These may include patient satisfaction, as well as adherence, compliance, staying on medications and side effects experienced. In routine care, physicians also have the task of providing cost-effective care, taking into account the costs and benefits of the therapeutic agents used. Against this background, a wider range of outcomes must be used in routine clinical care if well-founded therapy decisions are to be made and a high quality of care is to be ensured. In order to maintain comparability, it makes sense to use standards for outcome measurements in routine care as well. Ideally, even endpoints from clinical research can be adopted if they are practicable under everyday conditions. This facilitates the transfer of findings from clinical research to everyday conditions.

As an orientation for practice, the following parameters should be collected in HS standard care: (1) classification and diagnosis of the disease, (2) objective severity and (3) subjective severity / quality of life as outlined in the previous sections, (4) therapeutic needs, (5) patient preferences, and (6) documentation of comorbidity and comedication relevant for therapy.

In the treatment course, domains 1 and 6 are at best checked, but not used as therapeutic endpoints. In contrast, endpoints 2 to 5 are to be assessed as therapeutic benefits throughout the treatment course. In this context, safety and tolerability are added to the assessment of the therapeutic procedure. For the field of HS, no binding endpoints for routine care have yet been agreed upon. In some cases, only cursory data are documented under practice conditions; in more specialized centres, the endpoints of clinical studies are often documented as well. For the future, the establishment of an efficient standard data set under everyday conditions is desirable. The standard data set already created at the European level for registry research in HS is too extensive for this but should be a point of reference.

6 | MOVING FORWARD

Overall, novel, validated and reliably clinical endpoints, patient-reported outcomes, are being developed for both a clinical trial setting and for use in routine clinical practice.

In clinical trial settings, we are slowly moving away from the HiSCR as the primary outcome/assessment. The IHS4 and HASI-R are good contenders to measure treatment response. Regulatory authorities prefer dichotomous over continuous outcomes, and hence, the dichotomous IHS4 has been developed.11 Interestingly, a recent announcement by InflaRx mentioned that phase III vilobelimab study will use the FDA-suggested “modified HiSCR” as primary outcome.38 While it is said to include a reduction in all three lesion types (inflammatory nodules, abscesses and draining tunnels), exactly how this “modified HiSCR” will be calculated has not yet been made public.

Patient-reported outcomes are complementary to physician reported measures to provide a holistic assessment of HS disease severity within the core outcomes set being developed by the Hldradenitis Suppurativa cORe outcomes set International Collaboration (HISTORIC). The Hidradenitis Suppurativa Quality of
Life (HiSQOL) score is a validated measure of HS-specific quality of life and is already being used in many HS trials. An overall HS-specific quality of life question is also validated for use as a Patient Global Assessment. Magnitude of pain measurement via a 0–10 numerical rating scale is well-established; however, consensus is still required to ensure consistent administration and interpretation of the instrument.

It should be noted that although HS patients are adults, adolescent and rare paediatric cases also exist. Currently, there are no investigator-assessed or patient-reported outcomes measures specifically aimed at these patients.39,40

Feasibility of adopting the newly developed instruments is important and will ultimately influence their use in clinical registries and routine clinical practice going forwards.

AUTHOR CONTRIBUTIONS
Kelsey R. van Straalen, M. Augustin, Christos C. Zouboulis and John R. Ingram wrote a part of the manuscript, read and approved the final manuscript.

ACKNOWLEDGEMENTS
Thanks are owed to all the participants of the HISTORIC process, including patients, clinicians and methodologists. The Departments of Dermatology, Venereology, Allergology and Immunology, Dessau Medical Center, Dessau, Germany is a healthcare provider of the European Reference Network for Rare and Complex Skin Diseases (ERN Skin-ALLOCATE Skin group).

CONFLICT OF INTERESTS
All authors declare that none of the mentioned conflicts of interest had any influence to this manuscript. KRvS declares no conflicts of interest. JRI reports consultancy/advisory boards for Boehringer Ingelheim, ChemOcentrx, Inismed, Kymera Therapeutics, Novartis, UCB Pharma and Viela Bio. JRI is Treasurer of C3 (merger of CHORD & CS-COUSIN), HISTORIC Steering Committee member and lead for HISTORIC Pain domain work group. He is co-copyright holder of HiSQOL, Investigator Global Assessment and Patient Global Assessment for HS. JRI is BJD Editor-in-Chief and receives an authorship honorarium for two UpToDate HS chapters. MA reports consulting fees and/or research grants from AbbVie, Almirall, Amgen, Biogen, Boehringer Ingelheim, Celgene, Centocor, Eli Lilly, GSK, Hexal, Janssen, Leo Pharma, Medac, Merck, MSD, Mundipharma, Novartis, Pfizer, Sandoz, and UCB. CCZ reports consultancy/advisory boards disease-relevant honoraria from AbbVie, Bayer, Incyte, Inflarx, Janssen, Novartis, Regeneron, and UCB. He has received speaker fees from AbbVie and UCB. CCZ is President of the EHSF e.V., coordinator of the ALLOCATE Skin group of the ERN Skin and chair of the ARHS Task Force group of the EADV. He is Editor of the EADV News. CCZ is co-copyright holder of HIS4 on behalf of the EHSF e.V. His employer has received disease-relevant grants from AbbVie, Boehringer-Ingelheim, Inflarx, Novartis, and UCB for his participation as clinical investigator.

DATA AVAILABILITY STATEMENT
Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID
Kelsey R. van Straalen https://orcid.org/0000-0003-3305-3814
Matthias Augustin https://orcid.org/0000-0002-4026-8728
Christos C. Zouboulis https://orcid.org/0000-0003-1646-2608

REFERENCES


32. van Straalen KR. Chronic pain in hidradenitis suppurativa explained through the process of central sensitization. JAMA Dermatol. 2020;156:615-616.


