





# Development and validation of the International Hidradenitis Suppurativa Severity Score System (IHS<sub>4</sub>), a novel dynamic scoring system to assess HS severity

C.C. Zouboulis <sup>1,3</sup>, T. Tzellos,<sup>2</sup> A. Kyrgidis <sup>3</sup>, G.B.E. Jemec,<sup>4</sup> F.G. Bechara,<sup>5</sup> E.J. Giamarellos-Bourboulis,<sup>6</sup> J.R. Ingram <sup>7</sup>, T. Kanni,<sup>6</sup> I. Karagiannidis,<sup>1</sup> A. Martorell,<sup>8</sup> Ł. Matusiak,<sup>9</sup> A. Pinter,<sup>10</sup> E.P. Prens,<sup>11</sup> D. Presser,<sup>12</sup> S. Schneider-Burrus,<sup>13,14</sup> E. von Stebut,<sup>15</sup> J.C. Szepietowski,<sup>9</sup> H.H. van der Zee,<sup>11</sup> S.M. Wilden<sup>15</sup> and R. Sabat <sup>13</sup> on behalf of the European Hidradenitis Suppurativa Foundation Investigator Group\*

<sup>1</sup>Departments of Dermatology, Venereology, Allergology and Immunology, Dessau Medical Centre, Brandenburg Medical School Theodor Fontane, Dessau, Germany

<sup>2</sup>Department of Dermatology, Faculty of Health Sciences, University Hospital of North Norway, Harstad, Troms, Norway

<sup>3</sup>Division of Evidence Based Dermatology, Departments of Dermatology, Venereology, Allergology and Immunology, Dessau Medical Centre, Brandenburg Medical School Theodor Fontane, University of Brandenburg, Dessau, Germany

<sup>4</sup>Department of Dermatology, Zealand University Hospital, Roskilde, Denmark

<sup>5</sup>Department of Dermatology, Venereology and Allergology, Ruhr University Bochum, Bochum, Germany

<sup>6</sup>4th Department of Internal Medicine, Medical School, National and Kapodistrian University of Athens, Athens, Greece

<sup>7</sup>Department of Dermatology and Academic Wound Healing, Division of Infection and Immunity, Cardiff University, Cardiff, U.K.

<sup>8</sup>Department of Dermatology, Hospital de Manises, Valencia, Spain

<sup>9</sup>Department of Dermatology, Venereology and Allergology, Wroclaw Medical University, Wroclaw, Poland

<sup>10</sup>Department of Dermatology, Venereology and Allergology, Goethe University of Frankfurt, Frankfurt am Main, Germany

<sup>11</sup>Department of Dermatology, Erasmus University Medical Centre, Rotterdam, the Netherlands

<sup>12</sup>Department of Dermatology, Venereology and Allergology, University Hospital of Würzburg, Würzburg, Germany

<sup>13</sup>Interdisciplinary Group of Molecular Immunopathology, Dermatology/Medical Immunology, Charité – Universitätsmedizin Berlin, Berlin, Germany

<sup>14</sup>Centre of Vein and Skin Surgery, Havelkinik Berlin, Berlin, Germany

<sup>15</sup>Department of Dermatology, University Medical Centre, Johannes Gutenberg University Mainz, Mainz, Germany

## Summary

### Correspondence

Christos C. Zouboulis.

E-mail: christos.zouboulis@klinikum-dessau.de

### Accepted for publication

7 June 2017

### Funding sources

The development of the European Registry for Hidradenitis Suppurativa is supported by a research grant from the European Academy of Dermatology and Venereology (EADV-2015-023).

### Conflicts of interest

None declared.

C.C.Z., T.T. and A.K. contributed equally to this work.

\*See Appendix.

DOI 10.1111/bjd.15748

**Background** A validated tool for the dynamic severity assessment of hidradenitis suppurativa/acne inversa (HS) is lacking.

**Objectives** To develop and validate a novel dynamic scoring system to assess the severity of HS.

**Methods** A Delphi voting procedure was conducted among the members of the European Hidradenitis Suppurativa Foundation (EHSF) to achieve consensus towards an initial HS Severity Score System (HS<sub>4</sub>). Strengths and weaknesses of HS<sub>4</sub> were examined by a multicentre prospective study. Multivariate logistic regression, discriminant analysis and receiver operating characteristic curves, as well as examination for correlation (Spearman's rho) and agreement (Cohen's kappa) with existing scores, were engaged to recognize the variables for a new International HS<sub>4</sub> (IHS<sub>4</sub>) that was established by a second Delphi round.

**Results** Consensus HS<sub>4</sub> was based on number of skin lesions, number of skin areas involved and Dermatology Life Quality Index (DLQI), and was evaluated by a sample of 236 patients from 11 centres. Subsequently, a multivariate regression model calculated adjusted odds ratios for several clinical signs. Nodules, abscesses and draining tunnels resulted as the scoring variables. Three candidate scores were presented to the second Delphi round. The resulting IHS<sub>4</sub> score is arrived at by the number of nodules (multiplied by 1) plus the number of abscesses (multiplied by 2) plus the number of draining tunnels (multiplied by 4). A total score of 3 or less signifies mild, 4–10 signifies moderate and 11 or higher signifies severe disease. Cohen's kappa was fair ( $\kappa = 0.32$ ) compared with Hurley classification, and moderate ( $\kappa = 0.49$ ) compared with Expert Opinion.

Correlation was good ( $\rho > 0.6$ ) with Hurley classification, Expert Opinion, Physician's Global Assessment and Modified Sartorius score, and moderate for DLQI ( $\rho = 0.36$ ).

**Conclusions** The novel IHS4 is a validated tool to dynamically assess HS severity and can be used both in real-life and the clinical trials setting.

### What's already known about this topic?

- The modified Sartorius score, Hurley classification and Physician's Global Assessment have been used to assess severity of hidradenitis suppurativa.
- However, these are often either difficult to use in daily clinical practice or static and generally poorly validated.

### What does this study add?

- The proposed score is a systematically constructed, validated and simple tool to assess disease severity, and can be adapted both to clinical research and daily practice.

Hidradenitis suppurativa/acne inversa (HS) is a chronic, inflammatory, recurrent, debilitating skin disease.<sup>1,2</sup> It presents after puberty with painful, inflamed lesions in apocrine gland-bearing areas of the body, and leads to significantly impaired quality of life, depression and handicap. It exhibits many comorbidities, such as spondyloarthritis, inflammatory bowel disease, obesity and metabolic syndrome, which increase the disease burden.<sup>3–5</sup> Significant research efforts are being made to improve our understanding and management of this disease.

The increased research into the clinical aspects of HS highlights the need for a validated, easy to use disease severity assessment tool for HS that can be used both in clinical trials and daily clinical practice to accurately classify severity and guide therapeutic strategy.<sup>6</sup> Available clinical measures for assessing HS severity have recently been reviewed by Ingram *et al.*, and found unsatisfactory in several respects.<sup>7</sup> The systematic review included the Hurley classification, modified Sartorius score (MSS) and HS Physician's Global Assessment (HS-PGA).<sup>8–12</sup> All of these methods either exhibit weaknesses (such as no accurate assessment of extent of inflammation with each disease stage) or are described as time consuming and difficult to interpret.<sup>13</sup> Hidradenitis Suppurativa Clinical Response (HiSCR) was recently partially validated; however, it is designed to assess treatment response rather than disease severity cross-sectionally.<sup>14</sup>

The aim of this study was to establish a dynamic HS score. After an extensive first Delphi round by the European Hidradenitis Suppurativa Foundation (EHSF) members, consensus towards a Hidradenitis Suppurativa Severity Score System (HS4) was achieved.<sup>15</sup> The HS4 was evaluated in a prospective study and was optimized in order to demonstrate significant correlation with important parameters. Subsequently, a novel international HS4 (IHS4) was developed after a second Delphi round.

## Patients and methods

The STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) statement was used in the preparation of this manuscript.<sup>16</sup>

### Delphi voting procedures

A first Delphi voting procedure of five steps, leading to the establishment of HS4, was conducted among the EHSF members from 6 September to 19 October 2015 (Fig. 1). A two-thirds vote was preset as a requirement for agreement at steps 2 and 3, and a majority of votes for the last two steps. Of the 108 EHSF members on 6 September 2015, 69 (64%) participated in the procedure, and 42 of 69 participants contributed additionally by providing proposals. A second Delphi voting procedure was conducted from 5 to 11 December 2016 among the 42 EHSF members who had provided proposals leading to the decision on the IHS4 (Fig. 1). A majority of votes was required for agreement at this round.

### Patients

Validation of HS4 by correlation with other outcomes was performed through a prospective multicentre study at 11 centres in six different countries (Denmark, Germany, Greece, Netherlands, Poland and Spain). Inclusion of 20 patients was aimed at from each participating centre. All patients were diagnosed with HS using the Dessau diagnostic criteria.<sup>1</sup> Data, collected after patients provided written consent, were in compliance with the European registry platform, which is supported by the European Academy of Dermatology and Venereology.<sup>17</sup> The study protocol was approved by the Saxony-Anhalt Medical Association Ethics Committee (#20/16) as well as by ethics committees in the participating countries. Demographic (sex, age and centre) and

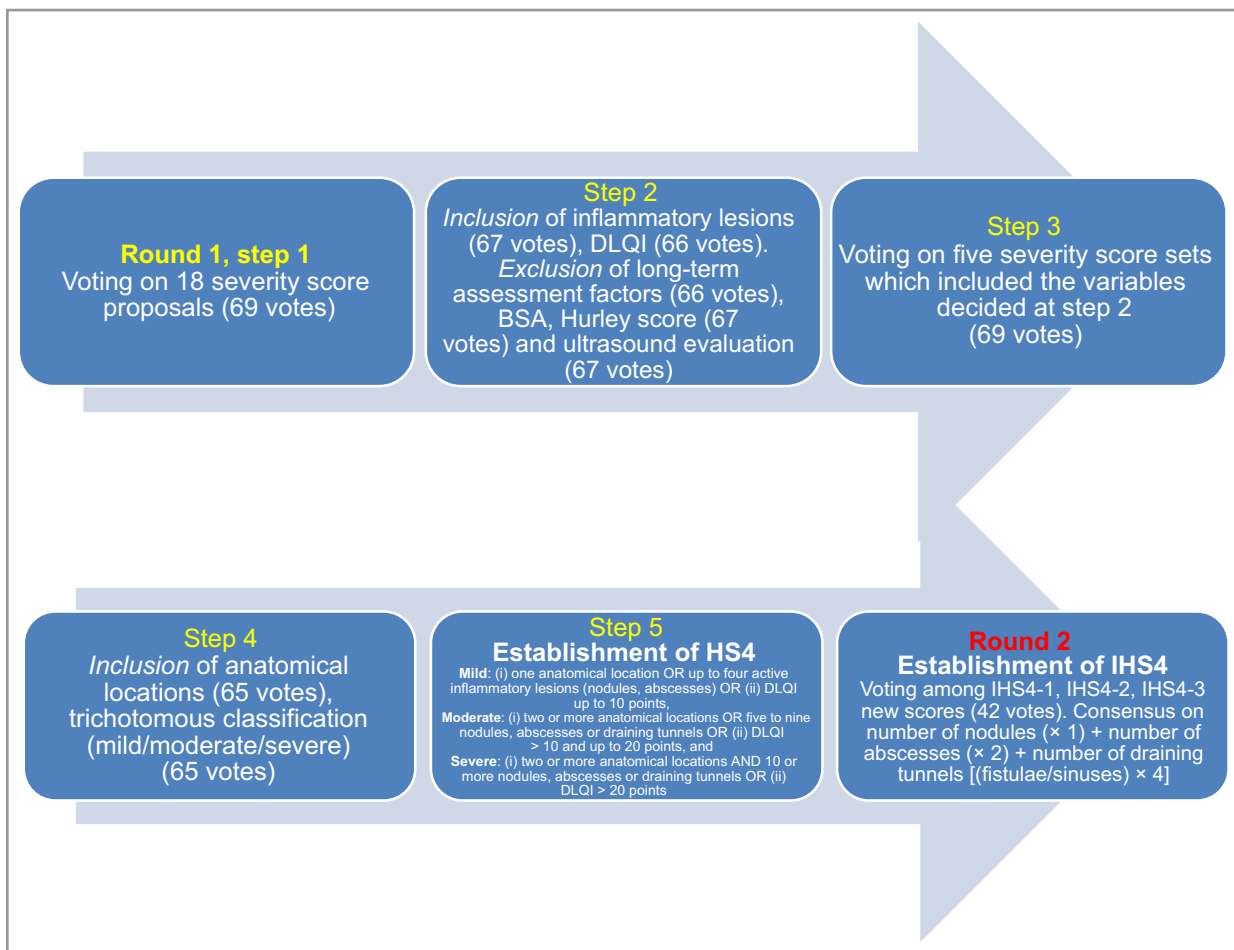


Fig 1. Flowchart of the Delphi voting procedure (two rounds, six steps in total). The resulting scores of steps 5 and 6 of the first and second rounds were validated through the multicentre dataset of 236 patients. DLQI, Dermatology Life Quality Index; BSA, body surface area; HS4, Hidradenitis Suppurativa Severity Score System; IHS4, International HS4.

clinical [smoking, weight, height, body mass index (BMI), previous medications, history of surgical treatment, perigenital involvement, number of locations involved, nodules, pustules, abscesses, draining tunnels (fistulae/sinuses), ulcers, etc.] variables were collected. Four physician-rated HS disease severity assessments were used as criteria measures for validation: Hurley classification by means, MSS, HS-PGA and Expert Opinion classification. Dermatology Life Quality Index (DLQI) was selected as the patient-reported outcome. We hypothesized that the new HS4 scoring system should be better than the existing scoring methods and therefore would not correlate perfectly with them.

### Statistical analysis

A sample size of 210 patients was predetermined. Assuming three groups of patients (i.e. mild, moderate and severe for Hurley classification), where each group would serve as the in-study group and both remaining groups would serve as comparators with an equal distribution, in an alternating order, the 210 patients would be equivalent to 70 in-study subjects and 140 comparator subjects. Based on pilot studies, we assumed a failure rate (proportion of subjects exhibiting a certain

dichotomous characteristic) among comparators of 0.3 and a failure rate for in-study subjects of 0.5. Therefore, our study would be able to reject the null hypothesis that the failure rates for in-study and comparator subjects are equal with a probability (power) of 0.807. The type I error probability associated with this test of the null hypothesis is 0.05. An uncorrected  $\chi^2$ -test statistic would be used to evaluate the null hypothesis.

Absolute and relative frequencies for demographic and clinical variables and existing scores were obtained. All variables were checked for normality via the Kolmogorov–Smirnov test. Non-parametric tests were used where applicable. Collinearity among scores and other variables were assessed via a correlation matrix, using Spearman's rho. Linear regression was used to model correlations between different scores. Crude odds ratios (ORs), adjusted ORs and corresponding 95% confidence intervals (CIs) were calculated by univariate and conditional multivariate logistic regression, respectively. For logistic regression analysis, outcome dichotomous variables were set to each score grade separately. Adjusted ORs and corresponding 95% CIs were calculated by multivariate logistic regression (backward elimination according to likelihood criteria). Variables with missing data were excluded from the regression analysis. To assess the internal

validity and degree of overoptimism (calibration) in the logistic regression models, the bootstrap resampling technique was applied by fitting the logistic model to a bootstrap sample of 66% of subjects drawn from the original sample over 100 repetitions.<sup>18</sup> To reduce the risk of obtaining significant results by chance due to multiple testing, we also performed multinomial logistic regression.

The logistic regression-derived model was then used to predict the event probability for a dichotomous outcome variable. Given the sets of independent variables produced by multivariate logistic regression, attempts were made to find linear combinations of those variables that could best separate groups of cases (discriminant analysis). Both all-variable and stepwise models were manually utilized. Stepwise models proved more parsimonious. The leave-one-out method was employed for classification schemes. Discriminant functions were saved, then receiver operating characteristic curves were used to choose between competing classification schemes. Agreement between finally selected scores was further tested via Cohen's kappa coefficient.

The alpha level was set at 0.05, while an alpha level of 0.10 was used as a cut-off for variable removal in the automated model selection for multivariate logistic regression. The type I error probability associated with all tests in this study was set to 0.05.

Statistical analyses were performed using IBM SPSS Statistics 23.0 (IBM, Armonk, NY, U.S.A.).

## Results

### Delphi consensus for development of the Hidradenitis Suppurativa Severity Score System

The first step of the first Delphi round led to collection of 18 severity score proposals from 42 participants. The proposals were grouped according to similarities of their variables and then submitted to a second voting step, which led to the following decisions (Fig. 1): inclusion of inflammatory lesions only (55 vs. 20 votes) and the DLQI (45 vs. 21), and exclusion of long-term assessment factors (19 vs. 47), body surface area (1 vs. 66), Hurley classification (12 vs. 55) and ultrasound diagnosis (0 vs. 67). There was consensus for the following variables: mild vs. moderate with a cut-off of five inflammatory lesions (18 vs. 2), and DLQI cut-off at 10 (36 vs. 10). Regarding the number of involved areas (33 vs. 34) and classification of mild/moderate/severe vs. mild/moderate (37 vs. 30), no final decision could be made. The third step of a vote among five score sets, which included the variables mentioned above, left the following undecided: (i) inclusion of anatomical locations and (ii) trichotomous (mild/moderate/severe) or dichotomous (mild/moderate) classification. The required fourth step confirmed inclusion of anatomical locations (38 vs. 27) and the trichotomous classification (35 vs. 30). The fifth step decided on the final variables, named HS4:<sup>15</sup> mild: (i) one anatomical location involved **or** up to four nodules or abscesses **or** (ii) DLQI up to 10 points; moderate: (i) two or more anatomical locations **or** five to nine nodules, abscesses or draining tunnels (fistulae/

sinuses) **or** (ii) DLQI more than 10 and up to 20 points; and severe: (i) two or more anatomical locations involved **and** 10 or more nodules, abscesses or draining tunnels (fistulae/sinuses) **or** (ii) DLQI more than 20 points.

### Descriptive statistics of the prospective multicentre study

Overall, 236 patients with HS from 11 study centres were included (Table S1; see Supporting Information), and comprised the sample group. Of these, 143 (60.6%) were women and 93 (39.4%) men. Mean age ( $\pm$  SD) of women was  $37.7 \pm 11.5$  years, while for men it was  $39.1 \pm 11.6$  years (Student's *t* test,  $P = 0.370$ ). Mean BMI was  $29.5 \pm 6.3$  kg m<sup>-2</sup>. Active smoking at baseline was reported by 156 (66.1%) patients. Table 1 presents sample baseline characteristics. Expert evaluation was mild HS in 69 (29.2%) patients, moderate in 91 (38.6%) and severe in 76 (32.2%) patients. Hurley score was I in 55 (23.3%) patients, II in 116 (49.2%) and III in 65 (27.5%) (Table S2; see Supporting Information). The HS-PGA score was mild in 51 of 200 patients (25.5%) evaluated by this score, moderate in 77 (38.5%) patients, severe in 22 (11.0%) patients and very severe in 50 (25.0%) patients (Table S2). Mean modified Sartorius score was  $59.65 \pm 48.68$ . Baseline mean DLQI was  $12.39 \pm 7.52$ .

### Explorations and attempts to improve the Hidradenitis Suppurativa Severity Score System

The HS4 score resulting from the first Delphi round was tested against Hurley classification and Expert Opinion classification. Sensitivity analyses were performed including all clinical signs documented, based on the results of the first Delphi round. For this procedure, each variable was either excluded or included one at a time, Boolean operators were changed and the number of lesions was reduced or increased. The prediction of Hurley classification was generally better if pustules were excluded. The best approach was without locations in the calculation. The best HS4 results achieved were a moderate correlation with the Hurley classification ( $\rho = 0.478$ ,  $P < 0.001$ ) and a good correlation with Expert Opinion classification ( $\rho = 0.692$ ,  $P < 0.001$ ) (Table 2) and HS-PGA ( $\rho = 0.672$ ,  $P < 0.001$ ). This best version of HS4 (HS4 improved) was (Table 2): mild: (i) up to four nodules or abscesses and (ii) DLQI up to 10 points; moderate: (i) five to nine nodules, abscesses or draining tunnels (fistulae/sinuses) or (ii) DLQI more than 10 and less than 20 points; and severe: (i) 10 or more nodules, abscesses or draining tunnels (fistulae/sinuses) or (ii) DLQI of more than 20 points.

### Explorations and attempts to improve/suggest Hidradenitis Suppurativa Physician's Global Assessment score

The possibility of utilizing the HS-PGA score to predict Hurley classification and Expert Opinion classification was subsequently assessed. A number of sensitivity analyses were

Table 1 Characteristics of the hidradenitis suppurativa patient sample

Sex	F (n = 143)	M (n = 93)	Total (n = 236)	
	Mean ± SD	Mean ± SD	Mean ± SD	Median (IQR)
Age	37.7 ± 11.5	39.1 ± 11.6	38.3 ± 11.6	38.2 (17)
Weight (kg)	84.4 ± 20.1	93.5 ± 21.3	88.0 ± 21.0	88.0 (28)
Height (cm)	168 ± 7	179 ± 7	173 ± 9	173 (12)
BMI (kg m <sup>-2</sup> )	29.7 ± 6.4	29.1 ± 6.1	29.5 ± 6.3	29.5 (9)
Use of antibiotics, n (%)	41 (28.7)	27 (29.0)	68 (28.8)	
Anti-TNF medication, n (%)	16 (11.2)	14 (15.1)	30 (12.7)	
Any other medication, n (%)	25 (17.5)	15 (16.1)	40 (16.9)	
History of surgical therapy, n (%)	101 (70.6)	71 (76.3)	172 (72.9)	
Currently smoking, n (%)	89 (62.2)	67 (72.0)	156 (66.1)	
Number of cigarettes per day <sup>a</sup>	9 ± 9	13 ± 11	11 ± 10	11 (20)
Disease duration (years)	14.9 ± 11.2	13.5 ± 10.0	14.3 ± 10.7	
Number of locations involved	3.9 ± 1.9	3.5 ± 1.6	3.7 ± 1.8	2.7 (3.0)
Number of locations involved including nonclassical ones	4.3 ± 2.1	3.9 ± 1.8	4.2 ± 2.0	4.2 (3)
Perigenital involvement, n (%)	120 (83.9)	78 (83.9)	198 (83.9)	
Nodules	6.3 ± 10.9	5.8 ± 6.3	6.1 ± 9.3	6.1 (7.0)
Pustules	4.9 ± 9.9	2.9 ± 4.8	4.1 ± 8.4	4.1 (5.0)
Abscesses	0.6 ± 1.4	1.3 ± 2.4	0.9 ± 1.9	0.9 (1.0)
Draining tunnels (fistulae/sinuses)	1.9 ± 3.6	3.7 ± 5.6	2.6 ± 4.6	2.6 (4.0)

F, female; M, male; IQR, interquartile range; BMI, body mass index; TNF, tumour necrosis factor. <sup>a</sup>Data missing from one patient.

Table 2 Summary of discriminant analysis, classification results and Spearman's rho (ρ)

Score system	Points	Expert opinion classification		Hurley classification	
		% original grouped cases correctly classified (discriminant analysis)	ρ	% original grouped cases correctly classified (discriminant analysis)	ρ
HS4		51.3 (AUC = 0.78)	0.52	37.7 (AUC = 0.71)	0.34
HS4 improved		67.8 (AUC = 0.86)	0.69	43.4 (AUC = 0.78)	0.48
IHS4-1	1 + 2 + 3	64.8 (AUC = 0.83)	0.75	54.7 (AUC = 0.81)	0.63
IHS4-2	1 + 2 + 4	66.1 (AUC = 0.82)	0.74	53.8 (AUC = 0.81)	0.64
IHS4-3	1 + 1 + 4	66.1 (AUC = 0.82)	0.74	54.2 (AUC = 0.81)	0.64

1 + 2 + 3 signifies (number of nodules × 1) + (number of abscesses × 2) + [number of draining tunnels (fistulae/sinuses) × 3]. All scores in the same manner, with changing multipliers. In Hidradenitis Suppurativa Severity Score System (HS4)-1 score cases, cut-off points are: 1 (mild) for < 3; 2 (moderate) for 3–10; and 3 (severe) for > 11. In HS4-2 and HS4-3 score cases, cut-off points are: 1 (mild) for < 4; 2 (moderate) for 4–10; and 3 (severe) for > 11. AUC, area under the curve; IHS4, International HS4.

similarly performed, in an attempt to further improve the HS-PGA score. Despite the good correlation of HS-PGA with the Hurley and Expert Opinion classifications (Table S3; see Supporting Information), the real level of agreement was not satisfactory. This was due to the 4-point scale of the HS-PGA score, which allows for better linear fitting but fails to correctly classify the 3-point Hurley classification ( $\kappa = -0.078$ ) and the Expert Opinion classification ( $\kappa = -0.16$ ).

### Development of a novel scoring instrument for hidradenitis suppurativa

We decided to attempt creation of a completely new scoring tool, which would be subsequently put into a second Delphi

round among the 42 EHSF members who had provided proposals for the first Delphi round. We used univariate and conditional multivariate logistic regression with a number of clinical variables to predict each Hurley/Expert Opinion classification grade as the dependent outcome variable. Results from the multivariate adjusted models for Hurley and Expert Opinion classifications are reported in Tables S4 and S5 (see Supporting Information).

By examining the adjusted ORs (Tables S4 and S5) as well as the results of the multinomial regression analysis (Table S6; see Supporting Information), we came to a series of conclusions. (i) It must be emphasized that draining tunnels (fistulae/sinuses) are a strong negative predictor for mild HS (both Hurley and Expert Opinion classifications). (ii) Nodules,

abscesses and draining tunnels (fistulae/sinuses) are strong positive predictors for severe grade HS (both Hurley and Expert Opinion classifications). (iii) Number of locations and perigenital involvement do not appear to have a role in Hurley or Expert Opinion classifications of patients.

Given the adjusted models above, nodules, abscesses and draining tunnels (fistulae/sinuses) were selected as variables to create the novel scoring system. As DLQI was found to limit the performance of the HS4, it was excluded from processing establishment of the new score.

A large number of scoring systems were created using the generic equation  $X + Y + Z$ , where X represents the number of nodules, Y the number of abscesses and Z the number of draining tunnels (fistulae/sinuses). We aimed for 3-point scores, although 4-point scores were also considered. Through a number of discriminant analyses, sensitivity analyses and manual and semiautomated explorations, the most suitable score systems were determined. Table 2 and Table S3 (see Supporting Information) present comparisons of the proposed score systems.

From the discriminant classification tables, score systems using 3 as a cut-off (IHS4-1) were found to be more effective in correctly classifying moderate cases. Scores using 4 as a cut-off (IHS4-2 and IHS4-3) were more efficient in correctly classifying mild cases. All scores shared the same efficacy in correctly classifying severe cases. IHS4-1, although achieving higher correlation, incorrectly classified a larger number of cases. All proposed scores correlated highly between each other. Because any of the proposed scores would perform similarly in the clinical setting, the second Delphi round was initiated, as described above, in order to choose the score with the best convenience and clinical meaningfulness. Scores presented to participants for casting their vote are listed in Table 2.

Overall, 35 of the 42 eligible members voted in the procedure. The votes were nine for IHS4-1, 20 for IHS4-2 and six for IHS4-3. Thus, EHSF members indicated that IHS4-2 is the most appropriate score to implement. Therefore, the proposed new International HS4 (IHS4) is:  $IHS4 \text{ (points)} = (\text{number of nodules multiplied by } 1) + (\text{number of abscesses multiplied by } 2) + [\text{number of draining tunnels (fistulae/sinuses) multiplied by } 4]$ . A score of 3 or less signifies mild, a score of 4–10 signifies moderate and a score of 11 or higher signifies severe HS (Fig. 2).

Cohen's kappa was fair ( $\kappa = 0.32$ ) for Hurley classification and moderate ( $\kappa = 0.49$ ) for Expert Opinion classification (weighted kappas were higher) (Table S7; see Supporting Information). Correlation was good ( $\rho > 0.6$ ) with all scores (Hurley classification, Expert Opinion classification, HS-PGA and MSS) (Table S3; see Supporting Information) and even for DLQI a moderate, significant correlation was reached ( $\rho = 0.36$ ,  $P < 0.001$ ). The correlation matrix of all HS current and newly developed scores is presented in Table S3 (see Supporting Information).

## Discussion

Accurate and reliable disease severity assessment and classification of HS patients are of utmost importance for guiding

<b>IHS4 (points) =</b>	
number of nodules	× 1 +
number of abscesses	× 2 +
number of draining tunnels (fistulae/sinuses)	× 4
<b>Mild HS:</b>	≤ 3 points
<b>Moderate HS:</b>	4–10 points
<b>Severe HS:</b>	≥ 11 points

Fig 2. International Hidradenitis Suppurativa Severity Score System (IHS4). A nodule (inflammatory nodule) is a raised, three-dimensional, round, infiltrated lesion with a diameter of >10 mm. An abscess is a tender but fluctuating mass with a diameter of >10 mm, surrounded by an erythematous area; the middle of an abscess contains pus. A draining tunnel is a raised, tender but fluctuating longitudinal mass of variable length and depth, ending at the skin surface, and sometimes oozing a fluid. Fistulae and sinuses are examples of tunnels.

therapeutic decisions as well for patient stratification in clinical studies.

We have established IHS4, a dynamic HS score, which achieved consensus among EHSF members (Fig. 2). This IHS4 score (points) = (number of nodules multiplied by 1) + (number of abscesses multiplied by 2) + [number of draining tunnels (fistulae/sinuses) multiplied by 4]. A score of 3 or less signifies mild HS, a score of 4–10 signifies moderate HS and a score of 11 or higher signifies severe HS.

All types of lesions included in the IHS4 are palpable with inflammatory signs (Fig. 3). A nodule (inflammatory nodule) is a raised, three-dimensional, round, infiltrated lesion with a diameter of > 10 mm.<sup>19</sup> An abscess is a tender but fluctuating mass with a diameter of > 10 mm, surrounded by an erythematous area; the middle of an abscess contains pus. A draining tunnel is a raised, tender but fluctuating longitudinal mass of variable length and depth, ending at the skin surface, and sometimes oozing a fluid.<sup>19</sup> Tunnels are morphologically variable, can extend beyond the skin and are not always related to a skin structure. Fistulae and sinuses are examples of tunnels.

The IHS4 is simple to calculate (Fig. 3) and validated with the use of existing physician-derived outcomes (HS-PGA, Hurley classification, MSS or Expert Opinion classification) and a patient-reported outcome measure (DLQI).

One key element of the new proposed score is that it includes only clinical signs of HS [nodules, abscesses and draining tunnels (fistulae/sinuses)] included in HiSCR. Thus it can be used in a complementary manner and simultaneously with HiSCR, and both can be calculated easily in daily clinical practice and clinical trial settings. IHS4 identifies mild cases and differentiates them from moderate and severe ones, which allows early identification of moderate and severe cases. The new scoring system proposes that presence of a draining tunnel (fistula/sinus) is sufficient to classify a case as at least



**Fig 3.** International Hidradenitis Suppurativa Severity Score System (IHS4) application, indicating mild, moderate and severe hidradenitis suppurativa (HS). White circle/ellipse, nodule; yellow circle/ellipse, abscess; red square/rectangle, draining tunnel; green circle/ellipse, papule; blue circle/ellipse, pustule; black square/rectangle, nondraining tunnel.

moderate. This property is important as it allows early intervention with systemic treatment before the disease progresses to irreversible status, which includes wide development of scar tissue; it also provides the possibility of assessing the effects of surgery. Furthermore, the current study indicates that the number of localizations and scar tissue may not be important variables when classifying severity. A study limitation is that the new score is derived from a given dataset and is exclusively physician-based. The IHS4 is intended for physicians to use, and should therefore be supplemented with patients' core outcome measure sets in future studies.

The Hurley classification stratifies patients into three stages. It was originally designed for selection of the appropriate treatment modality in a certain body location (medical therapy for Hurley stage I, local surgery for Hurley stage II and wide surgical excision for Hurley stage III).<sup>11,13</sup> However, it is static and was not designed as a dynamic score for accurate assessment of the extent of inflammation within each stage. MSS can be time consuming and difficult to interpret.<sup>13</sup> A new scoring system was recently suggested, but the variables included were determined by the authors, did not result from a regression analysis and the sample size was not justified and probably not sufficient.<sup>20</sup>

Similar to other skin disorders, such as psoriasis and atopic dermatitis, where disease severity is assessed, respectively, by the Psoriasis Area and Severity Index and the Scoring Atopic Dermatitis measure in both clinical trials and real life practice, the IHS4 is proposed as a valid, dynamic, simple and fast measurement.<sup>21,22</sup> Future studies should focus on validating IHS4 in other datasets as well. Furthermore, the fact that a Visual Analogue Scale (VAS) or other measure of pain was not included as a patient-derived outcome may be criticized.<sup>23,24</sup> Future studies assessing the validity of the proposed scoring system should also include this variable. We further consider a measure of patient quality of life (DLQI, VAS) to be essential as an additional set of criteria in globally evaluating a patient's condition.

In summary, IHS4 is a valid clinical scoring system for dynamic assessment of HS severity. Determining IHS4 requires counting nodules, abscesses and draining fistulas/sinus tracts, making it straightforward to apply in both research and clinical practice, and easy to use in conjunction with the HiSCR.

## References

- Zouboulis CC, Del Marmol V, Mrowietz U *et al.* Hidradenitis suppurativa/acne inversa: criteria for diagnosis, severity assessment, classification and disease evaluation. *Dermatology* 2015; **231**:184–90.
- Kurzen H, Kurokawa I, Jemec GB *et al.* What causes hidradenitis suppurativa? *Exp Dermatol* 2008; **17**:455–72.
- Fimmel S, Zouboulis CC. Comorbidities of hidradenitis suppurativa (acne inversa). *Dermatoendocrinol* 2010; **2**:9–16.
- Tzellos T, Zouboulis CC, Gulliver W *et al.* Cardiovascular disease risk factors in patients with hidradenitis suppurativa: a systematic review and meta-analysis of observational studies. *Br J Dermatol* 2015; **173**:1142–55.
- Gulliver W, Zouboulis CC, Prens E *et al.* Evidence-based approach to the treatment of hidradenitis suppurativa/acne inversa, based on the European guidelines for hidradenitis suppurativa. *Rev Endocr Metab Disord* 2016; **17**:343–51.
- 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**:619–44.
- Ingram JR, Hadjieconomou S, Piguet V. Development of core outcome sets in hidradenitis suppurativa: systematic review of outcome measure instruments to inform the process. *Br J Dermatol* 2016; **175**:263–72.
- Sartorius K, Killasli H, Heilborn J *et al.* Interobserver variability of clinical scores in hidradenitis suppurativa is low. *Br J Dermatol* 2010; **162**:1261–8.
- 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**:831–9.
- 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**:211–33.
- Hurley HJ. Hidradenitis suppurativa. In: Roenigk & Roenigk's *Dermatologic Surgery: Principles and Practice* (Roenigk RK, Roenigk HH Jr, eds), 2nd edn. New York: Marcel Dekker, Inc., 1996; 623–45.
- Kimball AB, Kerdel F, Adams D *et al.* Adalimumab for the treatment of moderate to severe hidradenitis suppurativa: a parallel randomized trial. *Ann Intern Med* 2012; **157**:846–55.
- van der Zee HH, Jemec GB. New insights into the diagnosis of hidradenitis suppurativa: clinical presentations and phenotypes. *J Am Acad Dermatol* 2015; **73** (Suppl. 1):S23–6.
- 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**:1434–42.
- Zouboulis CC, Alavi A, Bettoli V *et al.* Hidradenitis suppurativa severity score system (HS4): preliminary results. *Exp Dermatol* 2016; **25** (Suppl. 2):5.
- Hollestein LM, Nijsten T. Guidelines for statistical reporting in the *British Journal of Dermatology*. *Br J Dermatol* 2015; **173**:3–5.
- Daxhelet M, Suppa M, Benhadou F *et al.* Establishment of a European Registry for hidradenitis suppurativa/acne inversa by using an open source software. *J Eur Acad Dermatol Venereol* 2016; **30**:1424–6.
- Henderson AR. The bootstrap: a technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. *Clin Chim Acta* 2005; **359**:1–26.
- Lipsker D, Severac F, Freysz M *et al.* The ABC of hidradenitis suppurativa: a validated glossary on how to name lesions. *Dermatology* 2016; **232**:137–42.
- Chiricozzi A, Faleri S, Franceschini C *et al.* AISI: a new disease severity assessment tool for hidradenitis suppurativa. *Wounds* 2015; **27**:258–64.
- Fredriksson T, Pettersson U. Severe psoriasis – oral therapy with a new retinoid. *Dermatologica* 1978; **157**:238–44.
- Stalder JF, Taieb A, Atherton DJ *et al.* Severity scoring of atopic dermatitis: the SCORAD index. Consensus report of the European Task Force on Atopic Dermatitis. *Dermatology* 1993; **186**:23–31.
- Ring HC, Sørensen H, Miller IM *et al.* Pain in hidradenitis suppurativa: a pilot study. *Acta Derm Venereol* 2016; **96**:554–6.
- Zarchi K, Yazdanyar N, Yazdanyar S *et al.* Pain and inflammation in hidradenitis suppurativa correspond to morphological changes identified by high-frequency ultrasound. *J Eur Acad Dermatol Venereol* 2015; **29**:527–32.

## Appendix

Members of the European Hidradenitis Suppurativa Foundation Investigator Group include the following:

In Australia: R. Sinclair, D. Vekic; in Belgium: P. de Haes, V. del Marmol; in Canada: A. Alavi, W. Gulliver; in Denmark: G.B.E. Jemec; in France: A. Nassif; in Germany: F.G. Bechara, I. Karagiannidis, A. Kyrgidis, G. Nikolakis, A. Pinter, M. Podda, D. Presser, R. Sabat, S. Schneider-Burrus, E. von Stebut, S.M. Wilden, C.C. Zouboulis; in Greece: C. Antoniou, F.S. Delli, C. Dessinioti, E.J. Giamarellos-Bourboulis, T. Kanni, A. Patsatsi, A.P. Trigoni, C. Zisimou; in Italy: V. Bettoli, S. Garovich; in the Netherlands: J. Boer, B. Horváth, E.P. Prens, H.H. van der Zee; in Norway: G. Ingvarsson, T. Tzellos; in Poland: Ł. Matusiak, J.C. Szepietowski; in Spain: A. Martorell, L. Puig; in Switzerland: R.E. Hunger; in the U.K.: J.R. Ingram; in Sweden: L. Emtestam; in the U.S.A.: F.W. Danby (deceased), A.B. Kimball.

## Supporting Information

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

**Table S1.** Distribution of study participants in study centres.

**Table S2.** Baseline scoring classifications of the sample.

**Table S3.** Correlations of currently existing and proposed scores.

**Table S4.** Multivariate logistic regression. Predictors for Hurley classification (I, II or III) among study variables.

**Table S5.** Multivariate logistic regression. Predictors for Expert Opinion classification (mild, moderate or severe) among study variables.

**Table S6.** Multinomial logistic regression results. Adjusted predictors for mild, moderate and severe Hurley and Expert Opinion scores.

**Table S7.** Weighted Cohen's kappa coefficients for Hurley and Expert Opinion Scores vs. the International Hidradenitis Suppurativa Severity Score System-2.