

The refined Hurley classification: the inter-rater and intrarater reliability and face validity

DOI: 10.1111/bjd.18235

DEAR EDITOR, Hidradenitis suppurativa (HS) is a common, debilitating, chronic inflammatory skin disease, predominantly staged according to the Hurley classification. However, this classification was intended only to describe symptoms in one anatomical region and to guide surgical treatment options. With typically multiple areas affected by HS, this is not a valid instrument to classify the entire patient.¹ Because HS is a heterogeneous disease, the Dutch HS expert group proposed a modification, the 'refined Hurley classification'.² This consists of seven stages, subdividing stages I and II into stages A (mild), B (moderate) and C (severe), based on extent of the disease and degree of inflammation, while stage III (severe) is not subcategorized. This classification aims for a more detailed staging of patients with HS in daily practice and in clinical trials, ultimately to refine treatment strategies. The aim of this study was to assess the inter-rater and intrarater reliability and face validity of the refined Hurley classification.

A real-life assessment ($n = 25$) and a photographic assessment ($n = 15$) were performed in the Department of Dermatology, University Medical Center Groningen, The Netherlands, during the period May 2017 to July 2018. All adult patients with active HS visiting our clinic were eligible to participate. This real-life assessment consisted of two groups, each with two different independent raters. A fifth rater (B.H.) assessed all participants and this classification served as the reference. For the photographic assessment, participants were photographed according to a standardized protocol. All photographs were assessed by two independent investigators (L.M.P. and A.R.) for eligibility. At least two patients per refined Hurley stage were included.

A web-based survey was created using Qualtrics 2018 software (Provo, UT, U.S.A.) and was filled out twice, with an interval of 4 weeks, by 10 independent raters. All raters (12 residents and two dermatologists) received brief training on how to use the refined Hurley classification. Consulting the refined Hurley classification flowchart was permitted, as is possible in daily practice. Discussion between raters was not allowed. The study design followed the proposed Guidelines for Reporting Reliability an Agreement Studies.³ The inter-rater agreement was calculated as the percent agreement between raters. The Krippendorff alpha test with corresponding benchmarks was used to determine the inter-rater and intrarater reliability.⁴ Face




validity was evaluated by asking the raters to score the usefulness of the refined Hurley classification on a scale from 0 to 100.

For the real-life assessment, 25 patients were assessed: 13 in group 1 and 12 in group 2. The inter-rater agreement varied from 46.2% to 83.3% and the inter-rater reliability ranged from $\alpha = 0.68$ [95% confidence interval (CI) 0.32–0.95] to $\alpha = 0.92$ (95% CI 0.78–1.00). Compared with the reference classification, one rater (group 1) showed low inter-rater reliability ($\alpha = 0.60$; 95% CI 0.25–0.90), while the other three raters showed high inter-rater reliability: $\alpha = 0.88$ (95% CI 0.65–1.00) to 0.98 (95% CI 0.93–1.00).

In the photographic assessment, 86.7% of patients were identified as white with Fitzpatrick skin types I or II. The inter-rater reliability was $\alpha = 0.74$ (95% CI 0.71–0.78) for the first round and $\alpha = 0.80$ (95% CI 0.77–0.82) for the second round, while the intrarater reliability showed a mean α of 0.83 (95% CI 0.78–0.89). The inter-rater agreement for the refined Hurley stage for both time points is shown graphically in Figure 1. The face validity showed scores of 78.7 ± 10.3 prior the first photograph assessment and 76.5 ± 9.7 after the second assessment.

The original Hurley classification recently demonstrated a moderate inter-rater reliability and substantial intrarater reliability, based on photographic assessments.⁵ However, in our opinion, the original Hurley classification does not adequately reflect the disease extent and inflammatory activity of HS in the whole patient. For instance, patients with numerous widespread individual lesions (rated as refined Hurley 1C = severe), would still be classified as 'mild' in the original Hurley classification and consequently would not be eligible for treatment with biologics. A refinement of the original Hurley classification was therefore greatly needed. This is supported by a recent publication that showed an accurate correlation of the refined Hurley stages with HS severity assessed by both patients and clinicians.⁶ Other classification systems for HS, based on phenotypes, previously showed only low inter-rater reliability or had not yet been validated, leading to minimal use in daily practice.^{7,8}

In summary, the refined Hurley classification could be a reliable and useful tool for the classification and treatment of patients with HS in daily practice.

L.M. Prens ¹, A. Rondags,¹ R.J. Volkering,¹ I.C. Janse,² K. Politiek,¹ Y.S. Zuidema,¹ I. Turcan ¹, J. van Hasselt,¹ H.H. van der Zee³ and B. Horváth ¹

¹Department of Dermatology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands; ²Department of Dermatology,

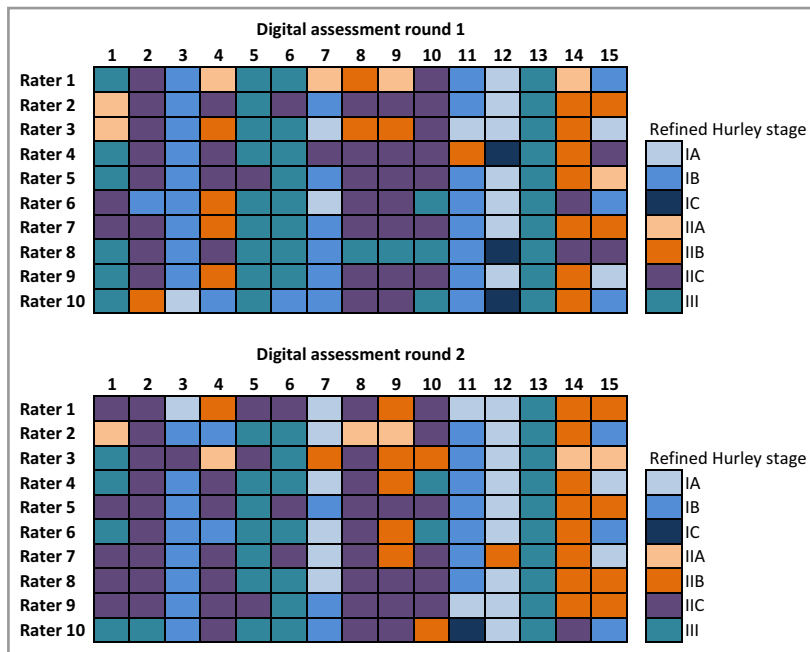


Fig. 1. Inter-rater agreement results of the photographic assessment.

Meander Medical Center, Amersfoort, the Netherlands; and ³Department of Dermatology, Erasmus Medical Center, Rotterdam, the Netherlands
E-mail: l.m.prens@umcg.nl

L.M.P. and A.R. contributed equally to this work and share joint authorship.

References

- Hurley HJ. Axillary hyperhidrosis, apocrine bromhidrosis, hidradenitis suppurativa, and familial benign pemphigus:surgical approach. In: *Dermatologic Surgery* (Roenigk RK, Roenigk HH, eds). New York: Marcel Dekker, 1989; 729–39.
- Horváth B, Janse IC, Blok JL et al. Hurley staging refined: a proposal by the Dutch Hidradenitis Suppurativa Expert Group. *Acta Derm Venereol* 2017; **97**:412–13.
- Kottner J, Audigé L, Brorson S et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011; **64**:96–106.
- Krippendorff K. *Content Analysis: An Introduction to Its Methodology*, 2nd edn. Beverly Hills: SAGE, 2004; 440.
- Ovadjia ZN, Schuit MM, van der Horst CMAM, Lapid O. Inter- and intrarater reliability of the Hurley staging for hidradenitis suppurativa. *Br J Dermatol* 2019; **181**:344–9.
- Rondags A, van Straalen KR, van Hasselt JR et al. Correlation of the refined Hurley classification for hidradenitis suppurativa with patient reported quality of life and objective disease severity assessment. *Br J Dermatol* 2018; **180**:1214–20.
- van Straalen KR, Verhagen T, Horváth B et al. Poor interrater reliability of hidradenitis suppurativa phenotypes. *J Am Acad Dermatol* 2018; **79**:577–8.
- van der Zee HH. New insights into the diagnosis of hidradenitis suppurativa: clinical presentations and phenotypes. *J Am Acad Dermatol* 2015; **73**:S23–6.

Funding sources: none.

Conflicts of interest: none to declare.