**Materials and Methods**

A total of 9 adult patients (3 females and 6 males) with confirmed diagnosis of HS of varying disease severity and symptom duration were included in the study. The patients were recruited at the Department of Dermatology, Venereology and Allergology of Wroclaw Medical University, Poland. All the subjects voluntarily agreed to join the study group and signed a written informed consent form. A group of 16 dermatology residents (15 females and 1 male) from the same department were invited to act as raters in the study. The physicians’ age ranged from 27 to 35 years, and their duration of dermatology experience was between 1 and 5 years (mean 2.7 ± 1.5 years). 7/16 raters were junior residents, during the first half of their residency, and 9/16 were senior residents, who had already finished at least 2.5 years of their residency or more. All of the residents had previously seen and dealt with some HS patients in a course of their practice in the hospital department or outpatient clinic. They had therefore the basic knowledge about disease manifestation, pathophysiology, and therapeutic methods. None of the raters had ever been involved in the regular treatment of any of the patients chosen for this study.

The scorings took place on a single day, when all subjects were assessed and scored by dermatology residents according to disease severity, using the Hurley Staging System, the Refined Hurley Staging, IHS4, HSSI, the Sartorius Score and HS-PGA. Also, the complete number of lesions, including nodules and draining tunnels, was counted. After being allocated into separated examination rooms and given numbers in order to identify the subjects, all the patients were examined twice by each physician with no time limit. The second assessments were performed directly after finishing the first ratings of all the patients. Physicians rotated between the rooms in a controlled order and recorded the data using standardized forms of scoring tools. They were also asked to work independently and not to contact each other until the study was completed. Additionally, just before the scorings, all the raters went through a short training session about using and interpreting the measurement tools and recording the data. The training was provided by an expert on HS, chosen from the department staff, the co-author of the paper (Ł.M.). During the training session, all the scoring systems were briefly characterized and discussed. The residents were also reminded how to recognize, distinguish, and name HS lesions.

Statistical analysis was performed using Statistica 12.0 software. Intrarater reliability was assessed using the intraclass correlation coefficient (ICC) with the following interpretation: poor reliability for ICC <0.40; moderate reliability 0.40–0.59; good reliability 0.60–0.74, and very good reliability for 0.75–1.00. The coefficient of variation (CV) was then calculated for the quantification of the interrater variability, and the ranges were as follows: 0–20% slight; 21–40% moderate; 41–60% high; >60% very high variability. After calculating means and standard deviations of the obtained values, correlations between all the investigated scales were verified using a Spearman’s rank order correlation test. The interpretation of Spearman’s correlation coefficient was as follows: 0–0.1 no correlation; 0.1–0.29 weak; 0.3–0.49 moderate; 0.5–0.7 strong; >0.7 very strong correlation. A Wilcoxon signed-rank test was used to analyze possible differences between both assessments. *p* values <0.05 were considered as significant.