ORIGINAL ARTICLE

■

Quantifying the natural variation in lesion counts over time in untreated hidradenitis suppurativa: Implications for outcome measures and trial design

John W. Frew, MBBS, MMed MS,[a] Caroline S. Jiang, MS,[b] Neha Singh, MS,[b] Kristina Navrazhina, BA,[c] Roger Vaughan, DrPH,[b] and James G. Krueger, MD, PhD[a]

New York, New York

Background: Hidradenitis suppurativa (HS) demonstrates high placebo response rates in clinical trials, possibly due to the natural variability of the disease. No quantification of variability in lesion counts of untreated disease has been undertaken.

Objective: To quantify the variability of untreated HS.

Methods: Deidentified individual patient data from the placebo arms of PIONEER studies were analyzed, and measurements of within-subject coefficients of variation were examined. Variability was stratified by disease-associated variables (Hurley stage, BMI, sex, smoking, family history) and body site.

Results: Analysis of within-subject coefficients of variation demonstrated that half of the participants had a middle spread [difference between 75th and 25th percentiles of the subject's abscess and nodule counts] greater than 33% and 40% of their median abscess and nodule counts, and 25% of the subjects had a middle spread greater than 70% and 78% of their median abscess and nodule counts in PIONEER I and II, respectively. Hurley stage 2 participants had significantly greater within-subject variation than Hurley stage 3 patients. Variation was greater in the axillary and groin regions than in other anatomical locations.

Limitations: Limitations include the use of precollected clinical trial data.

Conclusion: The within-subject variability of the lesion counts in untreated HS was greater than previously appreciated. This has profound effects on outcome measures and the conduct of future clinical trials of HS. (JAAD Int 2020;1:208-21.)

Key words: acne inversa; clinical trials; hidradenitis suppurativa; natural history; variability.

BACKGROUND

Hidradenitis suppurativa (HS) is a chronic inflammatory disease manifesting as painful nodules and abscesses and chronically draining tunnels in flexural areas.[1] Current outcome measures to assess disease severity in the clinic and in clinical trials rely upon counting the number of nodules, abscesses, and tunnels (also known as sinus tracts or fistulas).[2-4] The current gold standard outcome measure, HS clinical response (HiSCR),[5,6] is defined as a 50% reduction in the number of abscesses and nodules (AN count), with no increase in the number of abscesses or draining tunnels.[5,6] One known disadvantage of HiSCR is the elevated placebo response rates that occur with its use.[4,7] Placebo response rates in pivotal PIONEER studies of adalimumab for HS ranged between 26% and 27.6%, as measured by the percentage of participants achieving HiSCR.[7] Additionally, the placebo response rate in a phase 2b study of IFX-1 (a complement C5a) inhibitor, was very high at 47%,[4,8] with no statistically significant difference between the rates of HiSCR between the intervention and placebo arms.[4,8]

Factors that may contribute to the high placebo response rates include the sliding dichotomous nature of the outcome measure as well as the low inter-rater reliability of counting lesions.[2-4,9] A third aspect requiring further examination is the natural variation in lesion counts in untreated HS,[9] which remains incompletely defined.[4,9] Patient-reported retrospective questionnaires have been used to quantify the length of time for which the lesions persist (with 6.9 days reported for nodules).[10] However, given that clinical trials are based on an objective physician's evaluation of the clinical lesions,[7] the quantification of lesion variability needs to be assessed based on the physician's evaluation to maintain validity in the setting of clinical trials. A high level of test-retest reliability of the HiSCR outcome measure (R = 0.91) is at odds with the elevated placebo response rates previously described.[5] However, the primary endpoint for clinical trials is often at week 12 or 16,[7,11] suggesting that short-term observations may overestimate the stability of the lesion counts.

CAPSULE SUMMARY

- The natural variability of clinical disease in hidradenitis suppurativa remains incompletely defined.

- The within-subject variability of abscess and nodule counts in untreated hidradenitis suppurativa significantly contributes to placebo response rates and is greater for Hurley stage 2, axillary, and groin disease.

It has been acknowledged that individuals with AN counts of <3 at baseline should not be included in clinical trials as they would only require a one-lesion reduction (which may be in the range of natural disease fluctuation) in order to achieve HiSCR.[7] Additionally, such patients would not meet the criteria for moderate or severe disease. A number of studies have reported minimum AN counts of >5 in order to minimize issues with elevated placebo response rates[11]; however, the benefits are unclear given that the quantification of natural (untreated) disease variability has not been performed.[12]

The placebo arms of PIONEER I and II provided objective, physician-evaluated data for a period of 12 weeks of untreated (with the exception of topical antiseptic washes) moderate-to-severe HS. These data provide an opportunity to assess the natural variability of disease activity and develop a design for future clinical trials. Our overall aim was to assess the within-subject variability of lesion counts (inflammatory nodules, abscesses, draining tunnels, and AN counts) in participants enrolled in time period A (week 0 to week 12) in the placebo arms of the PIONEER I and PIONEER II phase 3 clinical trials using descriptive statistics and quantification of within-subject variations.

METHODS

Deidentified individual patient data from the PIONEER I and PIONEER II phase 3 studies of adalimumab therapy for HS were made available by AbbVie Inc and accessed through the secure Vivli online platform.[7] Raw data were extracted and compared with the available published data to ensure accuracy.[7] Data for participants allocated to the placebo arms in time period A (week 0 to week 12) were included in the analysis. Subjects were excluded if they had received any rescue antibiotics in PIONEER I or baseline concomitant antibiotics in PIONEER II. All data analyses were conducted using R version 3.5.3.[13]

Quantification of inflammatory nodules, abscesses, and draining tunnels (as previously defined) each available time point (Weeks 0, 2, 4, 6, 8, and 12) was undertaken.[5-7] As the lesion counts had skewed distributions (Fig 1), we examined 2 robust measures

of the within-subject coefficients of variation: (1) interquartile range (IQR) divided by the median (CVq) and (2) median absolute deviation (MAD) divided by the median (CVm). IQR is the difference between the 75th (Q3) and 25th percentile (Q1), the range of the middle half of the subject's data. Another measure of spread or dispersion that is more robust against outliers than IQR is (MAD). MAD is defined as the median of the absolute deviations from the subject's median value: MAD = median $(|x_i - \text{median}(x_i)|)$, where x_i represents the repeated measurements for a subject. Both measures of within-subject variation (CVq and CVm) were multiplied by 100 to obtain a percentage.

To illustrate this with an example, consider a subject with AN counts of 11, 12, 9, 8, and 7 from baseline to week 12. The median value of this subject's data is 9, with Q1 = 8 and Q3 = 11 (IQR = 3). The absolute deviations about the median are 2, 3, 0, 1, and 2, which in turn have a median value of 2. MAD for these data is 2, and the within-subject CVm is 22% (2/9 × 100). The midpoint of the absolute differences about the subject's median is 22% of their median AN count. The within-subject CVq is 33% (3/9 × 100) and is interpreted as follows: the spread of the middle half of the subject's data is 33% of their median AN count. Higher percentages of the within-subject coefficient of variation indicate more dispersion around the median and higher within-subject variability over the 12-week period. Note that the MAD is 0 when majority of the subject's lesion counts are the same. For example, a subject with abscess counts of 1, 1, 1, 4, and 5 has a MAD of 0 and CVm of 0% (0/1 × 100), whereas the CVq is calculated as 300% (3/1 × 100).

Within-subject variation was also visualized using spaghetti plots and heat maps. Stratification by the type of lesions (inflammatory nodule, abscess, draining tunnel, AN count), known *a priori* treatment efficacy-associated factors (Hurley stage, gender, BMI category, nicotine use, family history), and anatomical location was undertaken. The Mann−Whitney U test was used to compare the AN count's CVq and CVm values by Hurley stage, gender, nicotine use, and family history. The Kruskal−Wallis test was used to compare the AN count's CVq and CVm values by BMI category (underweight/normal [<25 kg/m^2], overweight [25 to <30 kg/m^2], and obese [≥30 kg/m^2], according to the Centers for Disease Control and World Health Organization guidelines for adults). $P < .05$ was considered statistically significant.

RESULTS

Of the 633 available participants identified in PIONEER I and II, 262 participants in the placebo arms were included in the analysis (n = 141 in PIONEER I and n = 121 in PIONEER II). The baseline demographic and disease characteristics of the included subjects are presented in Table I. Baseline disease activity was more severe in the PIONEER I subjects, with higher median lesion counts, IHS4 score, and proportion of subjects with Hurley stage 3, than in the PIONEER II subjects. The distributions of the baseline lesion counts for both the PIONEER studies are presented in Fig 1. The heat maps and spaghetti plots of the lesion counts of the inflammatory nodules, abscesses, draining fistulas, and AN counts over time for PIONEER II are presented in Figs 2 and 3. Additional plots for PIONEER I are presented in Figs 4 to 6.

The calculated measurements of the within-subject variance, stratified by inflammatory nodules, abscesses, draining tunnels, and AN counts, are presented in Table II. The distributions of these measures were non-normal (Fig 7). The median within-subject CVq ranged from 33% to 50% for the lesion counts in both PIONEER I and II. The median CVq for the AN count was 33% (Q1 = 14%, Q3 = 70%) in PIONEER I and 40% (Q1 = 19%, Q3 = 78%) in PIONEER II. The interpretation is that half of the participants had middle spreads (difference between the 75th and 25th percentiles of the subject's AN counts) greater than 33% and 40% of their median AN count in the placebo arms of PIONEER I and II, respectively. The 75th percentile of CVq for the AN count suggested that 25% of the subjects had middle spreads greater than 70% and 78% of their median AN counts in PIONEER I and II, respectively.

The spaghetti plots of the within-subject absolute deviations from the medians for PIONEER I and II are presented in Fig 8. The median within-subject CVm for the lesion counts ranged from 4% to 17% for PIONEER I and from 11% to 25% for PIONEER II. The within-subject variation was lower for abscesses than
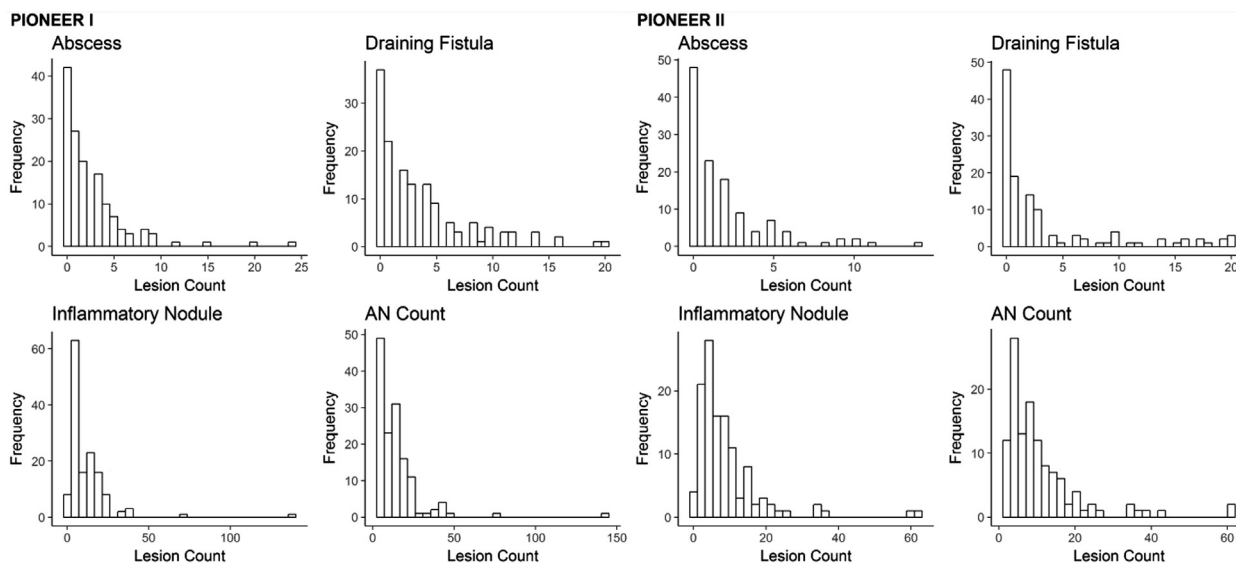
Fig 1. Histograms of the distribution of baseline disease activity in the PIONEER I and PIONEER II trials.

Table I. Baseline characteristics of the subjects included in the analysis

Characteristic	PIONEER I placebo (n = 141)	PIONEER II placebo (n = 121)
Gender		
Female	99 (70.2%)	90 (74.4%)
Male	42 (29.8%)	31 (25.6%)
Race		
White	110 (78.0%)	98 (81.0%)
Black	24 (17.0%)	15 (12.4%)
Other	7 (5.0%)	8 (6.6%)
Median Age*	36.0 (30.0-47.0)	33.0 (26.0-42.0)
Median BMI*	33.8 (28.4-39.3)	31.7 (26.5-36.5)
BMI category		
Underweight/normal	12 (8.5%)	22 (18.3%)
Overweight	35 (24.8%)	23 (19.2%)
Obese	94 (66.7%)	75 (62.5%)
Hurley stage		
2	75 (53.2%)	74 (61.2%)
3	66 (46.8%)	47 (38.8%)
Nicotine use	85 (60.3%)	89 (74.2)
Family history	28 (19.9%)	31 (25.6%)
Presence of draining tunnels	104 (73.8%)	73 (60.3%)
Median inflammatory nodules*	7 (4-15)	6 (4-10)
Median abscesses*	2 (0-3)	1 (0-3)
Median AN count*	11 (6-17)	8 (5-13)
Median draining tunnels*	2 (0-5)	1 (0-3)
Median baseline IHS4*	25 (13-40)	1 (9-30)

Data are presented as n values with percentages for categorical variables.
AN, Abscess and nodule.
*Data are presented as medians with interquartile ranges in parentheses.

for other lesion counts in PIONEER I when MAD was used as the measure of dispersion. Forty-three out of 87 subjects had the same abscess count for at least 3 out of the 5 visits, which resulted in a MAD of 0 and subsequently a CVm of 0%. The median CVm for the AN count was 16% (Q1 = 5%, Q3 = 31%) in PIONEER I and 20% (Q1 = 8%, Q3 = 39%) in PIONEER II. For half of the subjects, the MAD (middle of the absolute
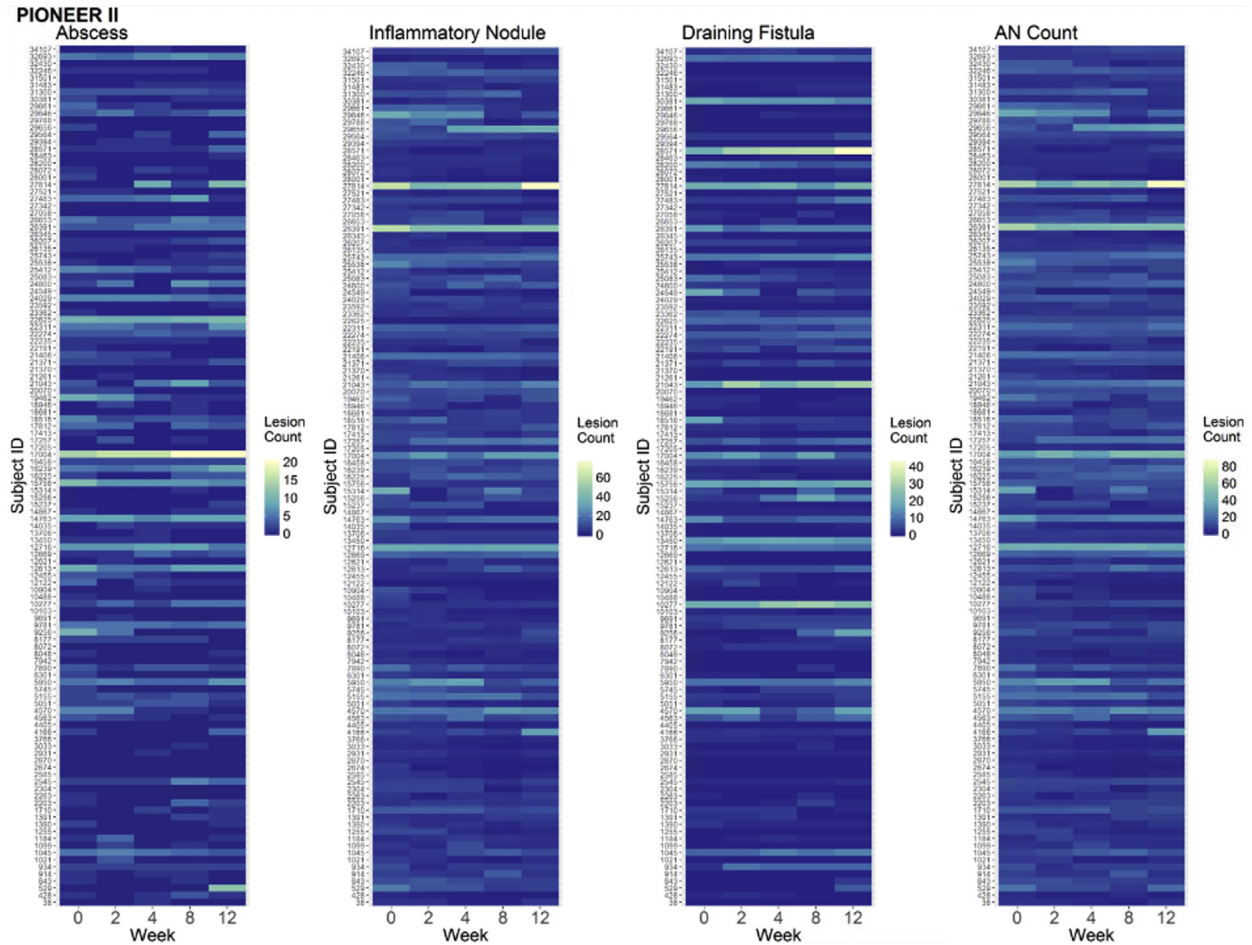
Fig 2. The heat maps of the lesion counts for the abscesses, inflammatory nodules, draining fistulas, and abscess and nodule (AN) counts during the 12-week time period of PIONEER II. These heat maps allow a visual representation of the within-subject variability for all the subjects included over time. The heat maps use a *blue-green-yellow* color scale, with the lowest count for each lesion type indicated in *blue* and the highest in *yellow*.

differences from the subject's median) was greater than 16% and 20% of their median AN counts in PIONEER I and II, respectively. For 25% of the subjects, the MAD was greater than 31% and 39% of their median AN counts in PIONEER I and II, respectively.

The small sample size for within-subject variation measures of the abscesses and draining tunnels in PIONEER I and II is noteworthy. The subjects in both the studies had low baseline counts for these lesions, and the coefficient of variation measures was undefined when divided by a median of 0. The within-subject variation measures of the AN count (sum of the number of abscesses and inflammatory nodule count important for determining the HiSCR response) were computable for 99.3% (140/141) of the subjects in PIONEER I and 97.5% (118/121) of the subjects in PIONEER II.

The stratification of the within-subject variation measures in AN counts by the *a priori* treatment efficacy-associated factors showed that the participants with Hurley stage 3 had significantly lower within-subject variability, as measured by CVq (Table III), compared to the participants with Hurley stage 2 in PIONEER I (P = .011) and II (P = .048). This difference was only observed in PIONEER I (P = .004) when CVm was used as the measure of within-subject variability (Table IV). No significant differences were observed when CVq and CVm were stratified by gender, BMI category, nicotine use, or family history.

The plots of the within-subject variations (measured by absolute deviations from the median AN counts) stratified by body site indicated that the axillary and inguinal regions demonstrated the greatest variability (Fig 9). The MAD was 0 across
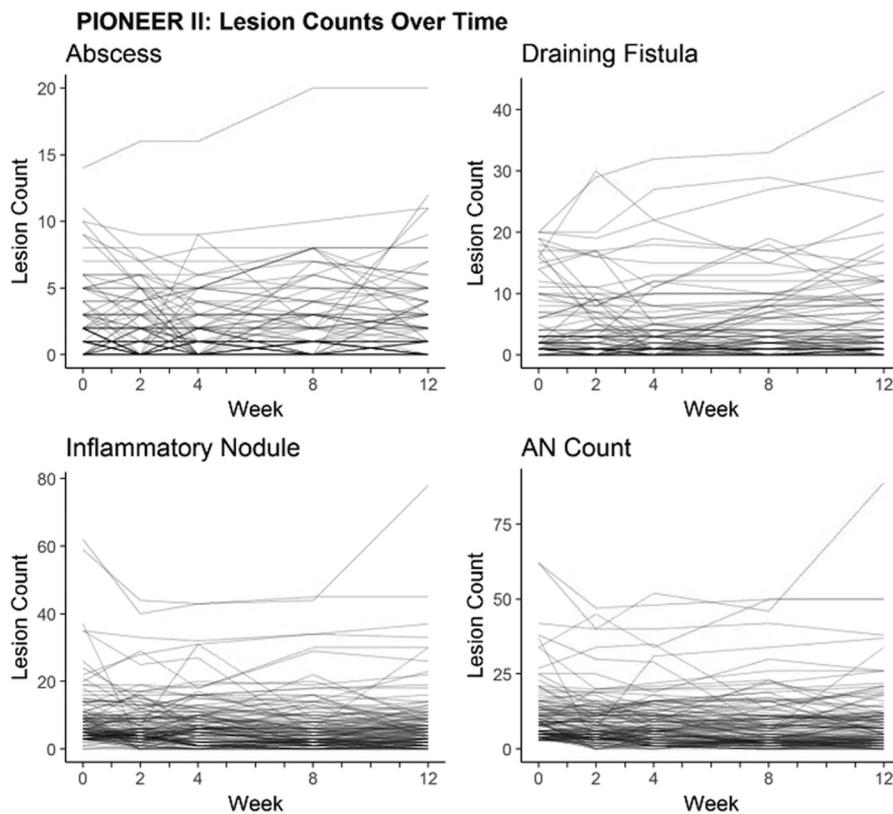
Fig 3. The spaghetti plots of the within-subject variability for abscesses, inflammatory nodules, draining fistulas, and abscess and nodule (AN) counts during the 12-week time period of PIONEER II. The x axis represents the absolute values of disease activity, and the y axis represents the time in weeks.

all body sites in PIONEER I and II. However, the means of their absolute deviations were consistently higher for the axillary and inguinal sites than for the other body sites (Table V). The measurements of CVq and CVm stratified by body site were not undertaken due to sparse AN counts when examining each body site separately. The CVq and CVm were calculated by dividing the median values by 0, which gave an undefined result across multiple body sites.

DISCUSSION

Our results have identified significant within-subject variability in the lesion counts in untreated participants from the placebo arms of the PIONEER I and II phase 3 studies. To our knowledge, this is the first report to quantify the natural variability of clinical HS and explain the high placebo response rates reported in the original studies (26.0% and 27.6%, respectively).[7] The median and IQR of the changes in the AN counts from the baseline (Table VI) and the placebo response rates were seen to progress over time (15.6% and 12.4% at week 2,

respectively; 20.1% and 24.0% at week 4, respectively; 22.5% and 28.1% at week 8, respectively; and 27.7% and 30.6% at week 12, respectively) in both PIONEER I and II. This supports the concept of natural variation in the disease, accounting for the results presented.

Participants in PIONEER II had a higher within-subject variation in the AN count compared to those in PIONEER I based on the quartiles (40% vs 33%, respectively) and MAD (20% vs 16%, respectively), which supports the higher placebo response rates seen in PIONEER II at weeks 4, 8, and 12 compared to those seen in PIONEER I. In PIONEER I, the Hurley stage 3 subjects had significantly lower variance in the AN count when compared to those with Hurley stage 2 when both the measures of the within-subject coefficient of variation were used. This may be due to the difficulty in assessing the abscesses and inflammatory nodules in the presence of significant scarring. The higher placebo response rates in PIONEER II may be explained by the higher proportion of Hurley stage 2 subjects than those in PIONEER I (Table I). This implies that individuals
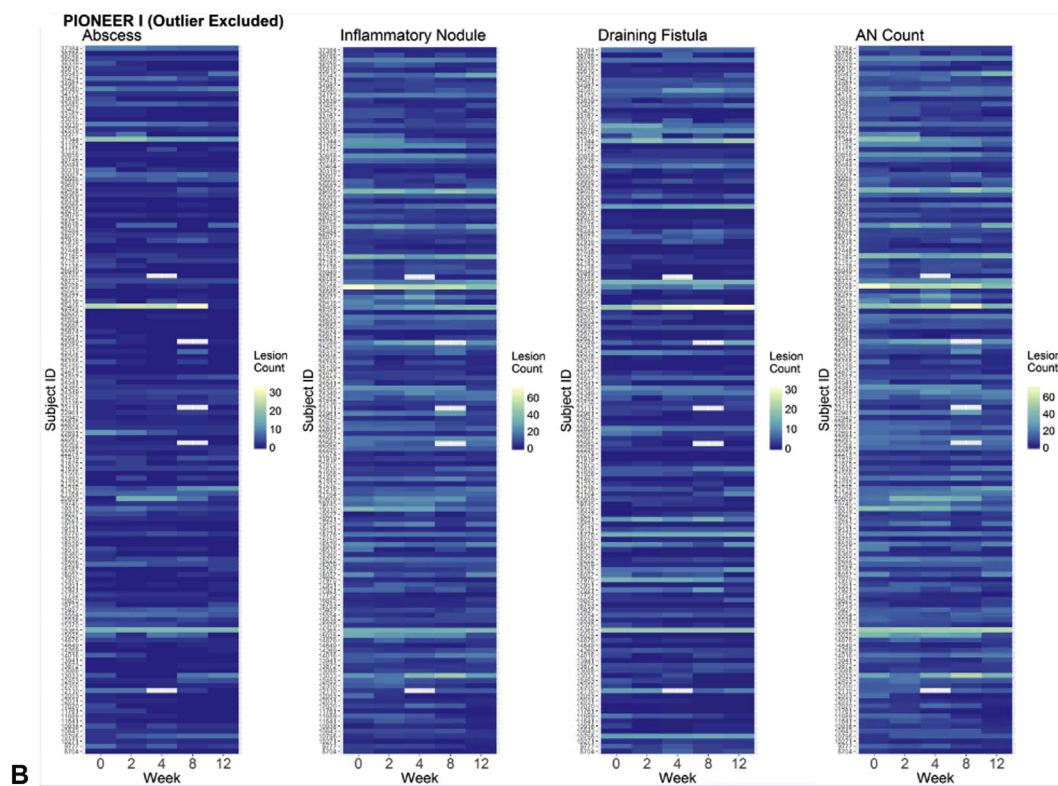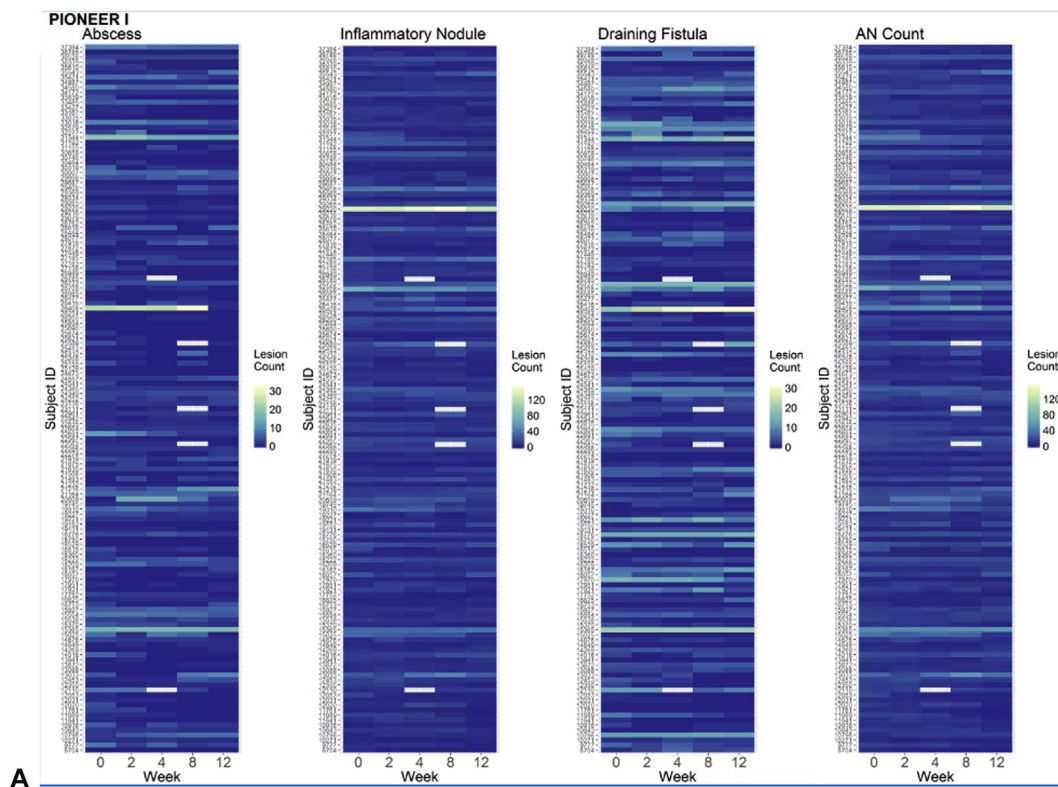
Fig 4. A and **B**, Heat maps of the lesion counts for the abscesses, inflammatory nodules, and draining fistulas and the abscess and nodule (AN) count during the 12-week time period of PIONEER I, with (A) and without (B) an outlier with consistently high inflammatory nodule counts throughout the 12-week period (this subject was included in all the analyses and excluded only here for visualization purposes).
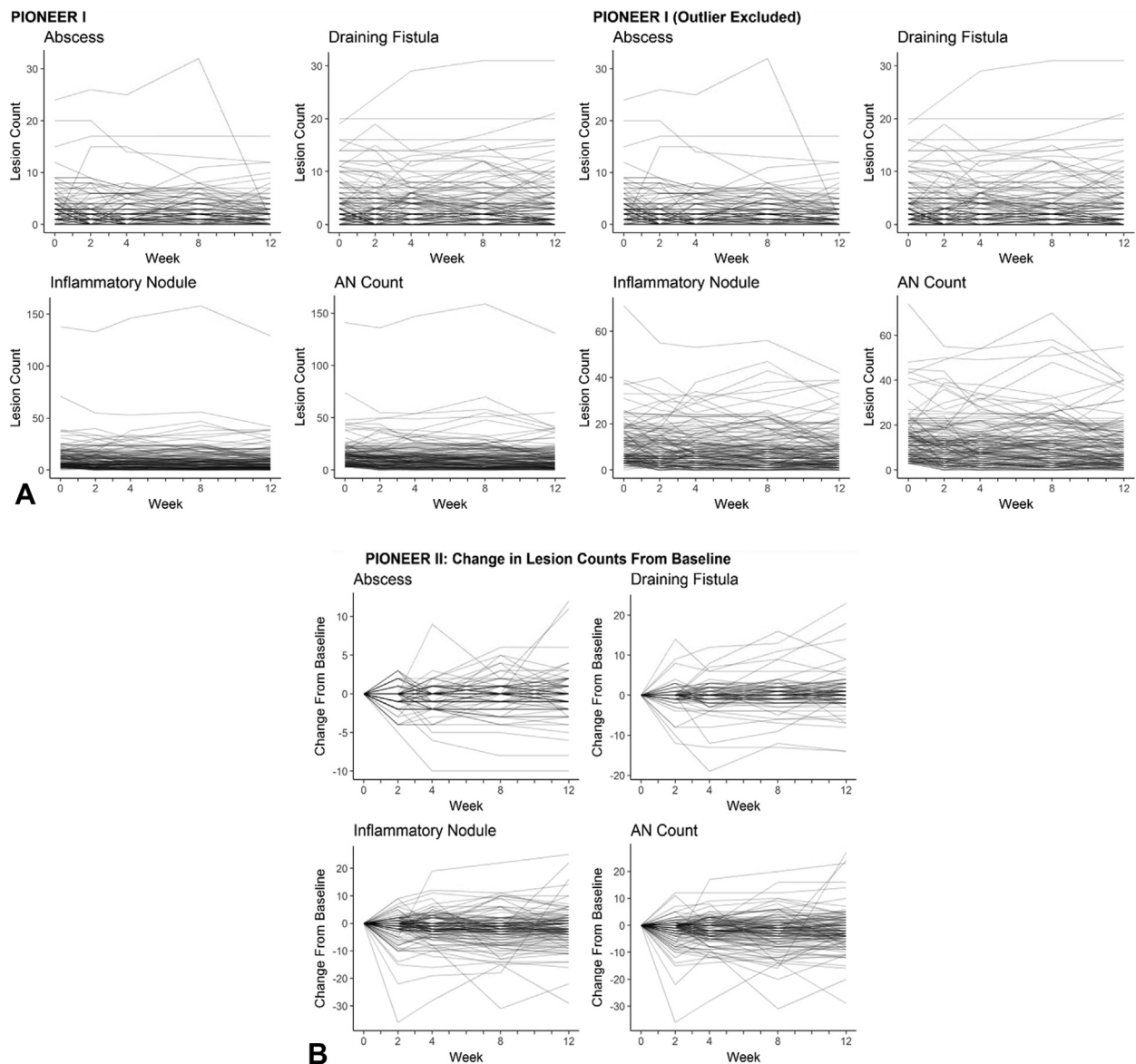
Fig 5. A, Spaghetti plots of the lesion counts for the abscesses, inflammatory nodules, draining fistulas, and abscess and nodule (AN) count during the 12-week time period of PIONEER I with and without an outlier with consistently high inflammatory nodule counts throughout the 12-week period (this subject was included in all analyses and excluded only for visualization purposes). **B**, Spaghetti plots of changes in the abscesses, inflammatory nodules, draining fistulas, and AN count during the 12-week time period of PIONEER II.

with lower disease severity are more prone to intra-subject variability.

Significant differences were identified when CVq was stratified by the morphological type of the lesion, with subepidermal lesions (abscesses/tunnels) having the greatest variation. Additionally, the within-subject variance was greater in the axillary and groin regions compared to that in the other anatomical locations. This variation between different anatomical sites may be inherent to the

disease and of such a degree that it does not carry any clinical significance; however, it must be acknowledged that "counting fatigue" by investigators in clinical trials may also play a role.

We can confirm that the analyzed data included coding regarding which investigator performed which ratings. For all individuals, the same investigator performed the ratings at the baseline and at all the time points through week 12. This indicates that the variability is due to intrapatient variation and not
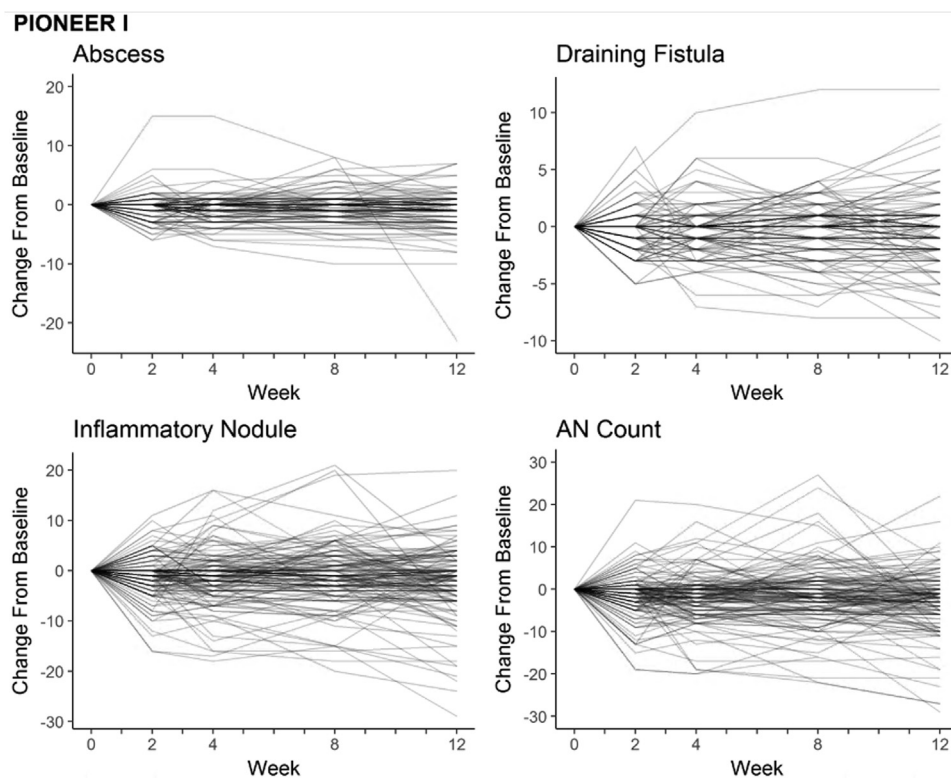
Fig 6. Spaghetti plots of change from baseline to each time point for abscess, inflammatory nodule, draining fistula, and AN count during period A (12 weeks) for PIONEER I.

Table II. The measures of within-subject variation in clinical disease activity of participants with hidradenitis suppurativa in the placebo arms of PIONEER I and II*

PIONEER I	Abscess (n = 87)	Inflammatory nodule (n = 137)	Draining fistula (n = 96)	AN count (n = 140)
Coefficient of variation quartiles (CVq) = IQR/ median × 100 (%)	50.0 (11.2, 100.0)	37.5 (18.8, 71.4)	40.0 (6.3, 100.0)	33.3 (14.0, 70.4)
Median absolute deviation (MAD) = median ($\mid x_i - $median$(x_i)\mid$)	0.0 (0.0, 1.0)	1.0 (0.0, 3.0)	0.0 (0.0, 1.0)	1.0 (1.0, 2.5)
Coefficient of variation MAD (CVm) = MAD/ median × 100 (%)	4.0 (0.0, 42.7)	16.7 (0.0, 37.5)	16.7 (0.0, 35.0)	15.8 (5.4, 31.4)
PIONEER II	**Abscess (n = 58)**	**Inflammatory nodule (n = 113)**	**Draining fistula (n = 73)**	**AN count (n = 118)**
Coefficient of variation quartiles (CVq) = IQR/ median × 100 (%)	45.0 (11.1, 100.0)	50.0 (25.0, 75.0)	33.3 (10.0, 100.0)	40.0 (18.7, 77.7)
Median absolute deviation (MAD) = median ($\mid x_i - $median$(x_i)\mid$)	0.0 (0.0, 1.0)	1.00 (1.00, 2.00)	0.0 (0.0, 1.0)	1.0 (1.0, 2.0)
Coefficient of variation MAD (CVm) = MAD/ median × 100 (%)	15.5 (0.0, 50.0)	25.0 (9.1, 44.4)	11.1 (0.0, 40.0)	20.0 (8.3, 39.4)

AN, Abscess and nodule; *IQR*, interquartile range (Q3-Q1); *MAD*, median absolute deviation; x_i, repeated measurements for a subject.
*Data are presented as medians with the 25th (Q1) and 75th (Q3) percentile values in parentheses. For the MAD calculations, n = 141 in PIONEER I and n = 121 in PIONEER II.

PIONEER I
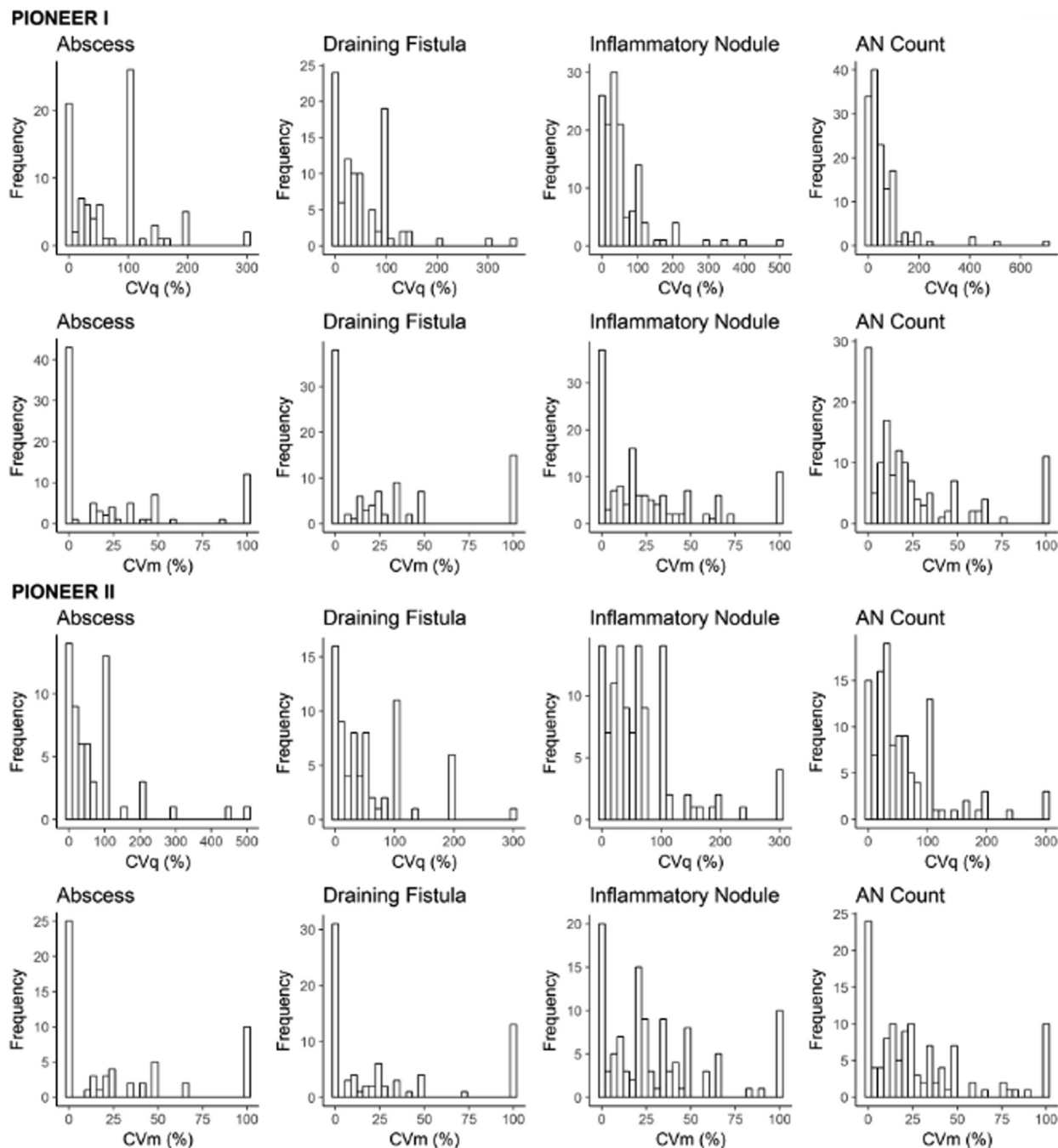
PIONEER II

Fig 7. Histograms of the distribution of within-subject variability measures in PIONEER I and PIONEER II.

inter-rater reliability. The progression of the placebo response rates over time would be more consistent with the intrasubject variability because of the natural variation in the disease activity rather than with the undocumented switching of the investigators. If there was any undocumented switching of investigators, this has not been captured in the dataset examined.

This analysis is limited by the inherent aspects of using clinical trial data as well as the restrictions according to the inclusion and exclusion criteria of the studies of interest. Other large databases of real-life clinical data exist, such as UNITE registry[14]; however, all of those patients were exposed to some form of systemic therapy. The placebo arms of the PIONEER studies provide the largest cohort of
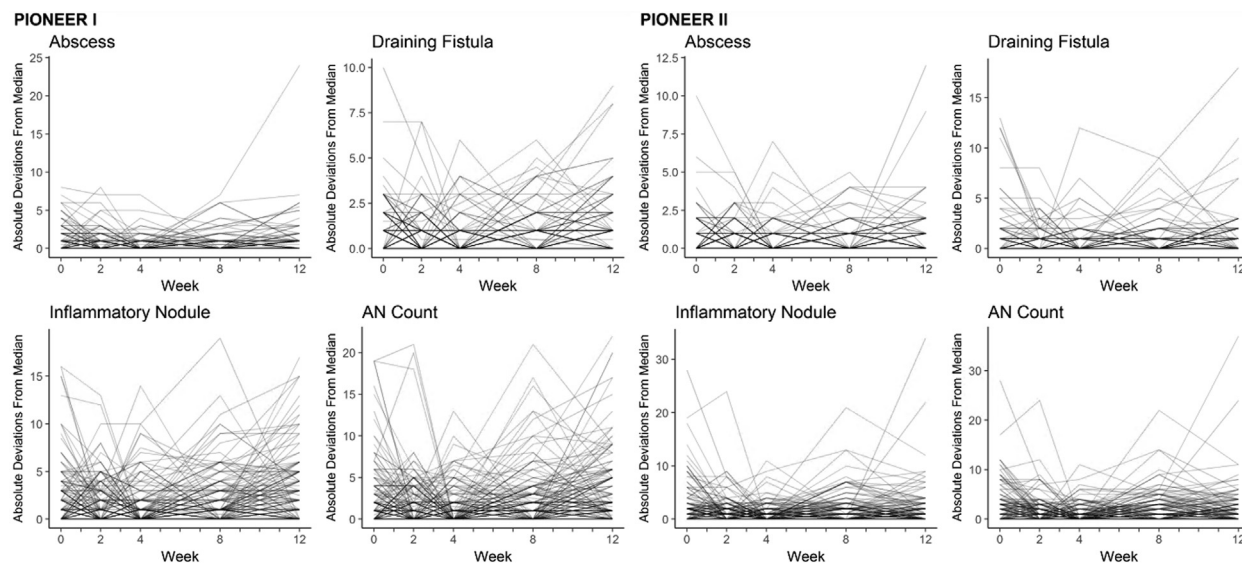
Fig 8. Spaghetti plots of the within-subject variability as measured by absolute deviations from the median for the abscesses, inflammatory nodules, draining fistulas, and abscess and nodule (AN) counts during the 12-week time period of PIONEER I and PIONEER II.

Table III. Within-subject variations in AN count as measured by coefficient of variation quartiles (cvq %) stratified by *a priori* features*

Characteristic	PIONEER I (n = 140)	*P* value	PIONEER II (n = 118)	*P* value
Gender				
Female	33.3 (14.9-72.0)	.684	40.0 (17.8-80.0)	.582
Male	32.1 (12.1-65.9)		33.3 (22.3-72.9)	
BMI category				
Underweight/normal	22.2 (13.6-55.7)		36.7 (17.8-100.0)	
Overweight	25.0 (11.1-71.3)	.607	64.5 (16.7-100.0)	.637
Obese	37.5 (16.7-70.0)		34.4 (21.8-66.7)	
Hurley stage				
2	48.5 (20.3-96.9)	**.011**	49.1 (26.0-100.0)	**.048**
3	23.8 (7.72-50.0)		33.3 (15.7-66.7)	
Nicotine use				
No	37.5 (14.6-66.2)	.939	34.4 (16.7-66.7)	.229
Yes	28.6 (13.2-80.0)		40.0 (20.0-83.7)	
Family history				
No	33.3 (14.8-72.5)	.518	35.9 (19.7-73.3)	.831
Yes	25.7 (11.1-58.8)		40.7 (16.7-96.9)	

AN, Abscess and nodule.
Bold numbers indicate statistical significance.
*Data are presented as medians with interquartile ranges in parentheses. Statistical analyses were conducted using Mann–Whitney U test for gender, Hurley stage, nicotine use, and family history and Kruskal–Wallis test for BMI category.

untreated patients for an extended time period, which is most applicable to investigate the placebo response rates in clinical trials. Further prospective studies of the natural histories of untreated patients are still needed and would further add to our knowledge in this aspect. The role of training investigators in improving the accuracy of HiSCR scoring and reducing variability remains unclear;

however, the investigators in the PIONEER studies underwent HiSCR training,[7] suggesting that the high degree of intrasubject variation is a true product of clinical disease activity. Our analysis was also limited by the inability to assess the natural variation in disease activity from potential confounding factors such as "counting fatigue," undocumented use of multiple investigators at a single site, and variations

Table IV. Within-subject variation in the AN count as measured by the coefficient of variation (cvm) based on the median absolute deviation stratified by *a priori* features*

Characteristic	PIONEER I (n = 140)	*P* value	PIONEER II (n = 118)	*P* value
Gender				
Female	16.23 (7.69, 32.69)		22.22 (10.00, 40.00)	
Male	14.36 (3.95, 30.22)	.527	13.39 (0.00, 31.25)	.123
BMI category				
Underweight/normal	11.11 (0.00, 36.36)		18.33 (0.00, 91.67)	
Overweight	14.29 (6.97, 34.29)		19.35 (10.42, 50.00)	
Obese	16.67 (5.26, 30.77)	.856	20.00 (9.32, 33.33)	.801
Hurley stage				
2	22.22 (10.00, 48.86)		22.65 (10.39, 40.00)	
3	11.11 (2.46, 20,00)	**.004**	14.84 (4.25, 30.78)	.12
Nicotine use				
No	16.67 (5.85, 28.57)		18.01 (5.63, 33.33)	
Yes	15.38 (5.00, 33.33)	.988	20.00 (8.33, 40.00)	.605
Family history				
No	16.23 (5.41, 31.41)		20.00 (7.50, 38.12)	
Yes	13.45 (4.69, 31.41)	.824	22.50 (8.75, 39.38)	.684

Bold numbers indicate statistical significance.
*Data are presented as medians with the 25th and 75th percentiles (Q1, Q3). Statistical tests were conducted using the Mann—Whitney U test for gender, Hurley stage, nicotine use, and family history and the Kruskal—Wallis test for BMI category.
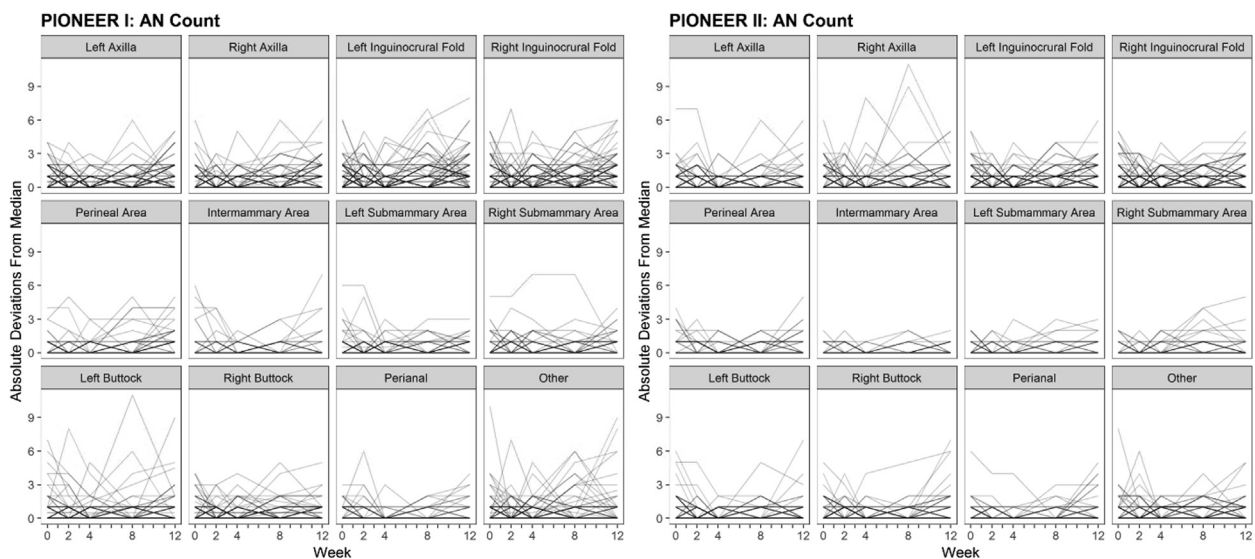
Fig 9. Spaghetti plots of the within-subject variability as measured by absolute deviations from the median for the abscess and nodule (AN) counts during the 12-week time period of PIONEER I and PIONEER II, stratified by body site.

in the counting methods between different investigators.

In order to address the identified within-subject variance and reduce the placebo response rates, various options could be employed in future HS clinical trials. Increasing the minimum number of lesions for inclusion in clinical trials would reduce the external validity of future trials and is not recommended. Elevating the clinical response definition from a 50% reduction to a >75%

reduction (>third quartile of CVq) in the AN count (HiSCR-75) would reduce the placebo response rates. Analyzing PIONEER II data based on HiSCR-75 would result in a placebo response rate of 15%, and 36% of those in the adalimumab arm would achieve HiSCR-75 ($P < .0001$). The third option would be to develop and validate a new clinical outcome measure that does not involve counting lesions and sliding dichotomous outcomes.

Table V. Stratification of the within-subject MAD stratified by body site*

Body site	PIONEER 1 (n = 141)		PIONEER 2 (n = 121)	
	Mean (min, max)	Median (Q1, Q3)	Mean (min, max)	Median (Q1, Q3)
Axilla				
Left axilla	0.22 (0.00, 2.00)	0.00 (0.00,0.00)	0.21 (0.00, 2.00)	0.00 (0.00, 0.00)
Right axilla	0.23 (0.00, 2.00)	0.00 (0.00, 0.00)	0.29 (0.00, 2.00)	0.00 (0.00, 0.00)
Groin				
Left inguinocrural fold	0.41 (0.00, 2.00)	0.00 (0.00, 1.00)	0.37 (0.00, 2.00)	0.00 (0.00, 1.00)
Right inguinocrural fold	0.40 (0.00, 3.00)	0.00 (0.00, 1.00)	0.36 (0.00, 2.00)	0.00 (0.00, 1.00)
Perineal area	0.13 (0.00, 3.00)	0.00 (0.00, 0.00)	0.10 (0.00, 2.00)	0.00 (0.00, 0.00)
Submammary				
Intermammary area	0.05 (0.00, 2.00)	0.00 (0.00, 0.00)	0.03 (0.00, 1.00)	0.00 (0.00, 0.00)
Left submammary area	0.09 (0.00, 2.00)	0.00 (0.00, 0.00)	0.07 (0.00, 2.00)	0.00 (0.00, 0.00)
Right submammary area	0.14 (0.00, 5.00)	0.00 (0.00, 0.00)	0.05 (0.00, 2.00)	0.00 (0.00, 0.00)
Buttock				
Left buttock	0.19 (0.00, 4.00)	0.00 (0.00, 0.00)	0.14 (0.00, 3,00)	0.00 (0.00, 0.00)
Right buttock	0.17 (0.00, 4.00)	0.00 (0.00, 0.00)	0.13 (0.00, 4.00)	0.00 (0.00, 0.00)
Perianal	0.06 (0.00, 1.00)	0.00 (0.00, 0.00)	0.08 (0.00, 4.00)	0.00 (0.00, 0.00)
Other	0.21 (0.00, 3.00)	0.00 (0.00, 0.00)	0.17 (0.00, 2.00)	0.00 (0.00, 0.00)

MAD, Median absolute deviation.
*Data are presented as mean with ranges (min, max) or as medians with 25th and 75th percentiles (Q1, Q3).

Table VI. Changes in the AN count at each time point from the baseline*

Time point	PIONEER I (n = 141)	P value	PIONEER II (n = 121)	P value
Week 2	−1 (−3, 0)	.0002	0 (−3, 0)	<.0001
Week 4	−1 (−4, 1)	<.0001	−2 (−4, 1)	<.0001
Week 8	−2 (−5, 2)	.001	−2 (−6, 1)	<.0001
Week 12	−2 (−6, 0)	<.0001	−2 (−5, 1)	<.0001

AN, Abscess and nodule.
*Data are presented as medians with 25th and 75th percentiles (Q1, Q3). *P*-values were calculated using the Wilcoxon signed-rank test.

CONCLUSIONS

The natural variability in the lesion counts in the untreated participants in the placebo arms of PIONEER I and II was significantly larger than previously appreciated. This variation was greater among Hurley stage 2 than stage 3 subjects and also varied by body site. Clearly, if lesion counts continue to be used as primary outcomes, this higher than previously appreciated variability will have an untoward impact on the required sample sizes, costs, and time to completion for future clinical trials.

This publication is based on research data from AbbVie Inc, which have been made available through Vivli Inc. Vivli Inc and AbbVie Inc had no role in the design or execution of the study, statistical analysis, or composition of the manuscript.

REFERENCES

1. Sabat R, Jemec GBE, Matusiak L, Kimball AB, Prens E, Wolk K. Hidradenitis suppurativa. *Nat Rev Dis Primers.* 2020;6(1):18.
2. Thorlacius L, Garg A, Riis PT, et al. Inter-rater agreement and reliability of outcome measurement instruments and staging systems used in hidradenitis suppurativa. *Br J Dermatol.* 2019; 181(3):483-491.
3. 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.
4. Frew JW, Jiang CS, Singh N, et al. Clinical response rates, placebo response rates, and significant associated covariates are dependent on choice of outcome measure in hidradenitis suppurativa. A post hoc analysis of PIONEER 1 and 2 individual patient data. *J Am Acad Dermatol.* 2020;82(5):1150-1157.
5. Kimball AB, Jemec GB, Yang M, Kalgeleriy A, Signorovith JE, Okun MM. 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.
6. Kimball AB, Sobell JM, Zouboulis CC, Gu Y, Williams DA, Sundaram M. HiSCR (Hidradenitis Suppurativa Clinical Response): a novel clinical endpoint to evaluate therapeutic outcomes in patients with hidradenitis suppurativa from the placebo controlled portion of a phase 2 adalimumab study. *J Eur Acad Dermatol Venereol.* 2016;30(6):989-994.
7. Kimball AB, Okun MM, Williams DA, et al. Two phase 3 trials of adalimumab for hidradenitis suppurativa. *N Engl J Med.* 2016; 375:422-434.
8. InflaRx. InflaRx announces top-line SHINE Phase IIb results for IFX-1 in hidradenitis suppurativa. Available at: https://www.inflarx.de/Home/Investors/Press-Releases/06-2019-InflaRx-Announces--Top-Line-SHINE-Phase-IIb-Results-for-IFX-1-in-Hidradenitis-Suppurativa-.html. Accessed September 10, 2020.
9. Ali AA, Seng EK, Alavi A, Lowes MA. Exploring changes in placebo treatment arms in hidradenitis suppurativa randomized clinical trials: a systematic review. *J Am Acad Dermatol.* 2020;82(1):45-53.

10. von der Werth JM, Williams HC. The natural history of hidradenitis suppurativa. *J Eur Acad Dermatol Venereol*. 2000; 14(5):389-392.

11. Maarouf M, Clark AK, Lee DE, Shi VY. Targeted treatments for hidradenitis suppurativa: a review of the current literature and ongoing clinical trials. *J Dermatolog Treat*. 2018;29(5): 441-449.

12. Bland JM, Altman DG. Measurement error proportional to the mean. *BMJ*. 1996;313(7049):106.

13. R Core Team. R: A language and environment for statistical computing. R Project for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/. Accessed November 3, 2020.

14. Prens EP, Lugo-Somolinos AM, Paller AS, et al. Baseline characteristics from UNITE: an observational, international, multicenter registry to evaluate hidradenitis suppurativa (acne inversa) in clinical practice. *Am J Clin Dermatol*. 2020;21(4):579-590.