Received: 5 January 2024 | Revised: 21 March 2024 | Accepted: 15 April 2024

DOI: 10.1002/jvc2.441

JEADV CLINICAL PRACTICE
OPEN ACCESS JOURNAL OF THE EUROPEAN ACADEMY OF DERMATOLOGY & VENEREOLOGY

ORIGINAL ARTICLE

Determining meaningful thresholds for evaluating treatment efficacy in patients with alopecia areata

Ernest H. Law[1] | **Nicole J. Williams**[2] | **Dane Korver**[2] | **Randall H. Bender**[2] | **Debanjali Mitra**[1] | **Gregor Schaefer**[3] | **Lauren M. Nelson**[2]

[1]Global Access & Value, Pfizer Inc., New York, New York, USA

[2]Patient-Centered Outcomes Assessment, RTI Health Solutions, Research Triangle Park, Research Triangle Park, North Carolina, USA

[3]Global Medical Affairs, Pfizer Pharma GmbH, Berlin, Germany

Correspondence
Ernest H. Law, Global Access & Value, Pfizer Inc., 235 East 42nd St, New York, NY 10017, USA.
Email: Ernest.Law@pfizer.com

Funding information
Pfizer

Abstract

Background: The Severity of Alopecia Tool (SALT) is a clinician-reported outcome measure of scalp hair loss in alopecia areata (AA).

Objectives: To characterise the magnitudes of change in SALT scores corresponding to meaningful treatment benefits from the patient's perspective.

Methods: Anchor-based methods for the estimation of meaningful within-patient change thresholds were applied to pooled data from a randomised, double-blind trial of ritlecitinib. Anchors included a patient-reported measure of change in AA severity, the Patient Global Impression of Change (PGI-C) and three items comprising the Patient Satisfaction with Hair Growth (P-Sat) questionnaire. After reviewing Pearson correlations between change-from-baseline SALT scores and each anchor to confirm adequate association, potential thresholds were computed as mean change-from-baseline SALT scores among patients who reported moderate improvement on the PGI-C and/or moderate satisfaction on each of three P-Sat items at week 24.

Results: Six hundred and fifty participants (86% adults, 14% adolescents) had mean (standard deviation) SALT scores of 90.6 (14.3) at baseline, suggesting a sample with primarily severe AA. Correlations between SALT change-from-baseline scores and the patient-reported items supported their use as anchors. Estimates based on patients reporting moderate improvement in AA ($n = 102$) on the PGI-C and those reporting moderate satisfaction on the P-Sat item related to the amount of hair growth at week 24 ($n = 122$) were −42.2 (26.1) and −43.1 (26.8), respectively. Supportive estimates based on the remaining P-Sat items were similar in magnitude.

Conclusions: Among patients with severe AA, SALT change-from-baseline scores of 42 or 43 represent meaningful improvements. While the achievement of low SALT scores of ≤10–≤20 have been used to characterise efficacy in clinical trials, the amount of change required to meet this endpoint far exceeds

the estimates in this study. The treatment goals of individual patients must be considered when evaluating benefit in both clinical trials and clinical practice.

KEYWORDS
alopecia areata, hair loss, meaningful change, patient-reported outcomes, SALT

INTRODUCTION

Alopecia areata (AA) is an autoimmune disease that targets hair follicles but does not destroy them. AA is characterised by nonscarring hair loss ranging from small bald patches to complete loss of hair on the scalp, face and/or body.[1] AA has a far-reaching impact on patients' quality of life and psychological health, including lowered self-esteem and increased incidence of anxiety and depressive disorders.[2–4] The treatment landscape for AA is evolving: baricitinib, a Janus kinase (JAK) inhibitor, has been approved by the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) for the treatment of adults (≥18 years) with severe AA. The efficacy and safety of the JAK inhibitor ritlecitinib was evaluated in the phase 2b/3 ALLEGRO-2b/3 trial (NCT03732807), and ritlecitinib has been approved by the FDA and the EMA for the treatment of severe AA in adults and adolescents 12 years and older.

The Severity of Alopecia Tool (SALT), used widely in clinical research, is administered according to guidelines published by the National Alopecia Areata Foundation for evaluating percentage hair loss in clinical trials of AA[5–7]: a patient's SALT score represents the degree of scalp hair loss due to AA, ranging from 0 (no scalp hair loss) to 100 (complete scalp hair loss).[5] The percentage area of hair loss in each of the four quadrants is multiplied by the proportion of total scalp surface area for the respective view and summed to give the SALT score (total percentage area of hair loss). Qualitative research in AA—which has provided an initial, informative investigation into treatment benefit—suggests that most patients would consider a treatment successful if scalp hair loss of 50% or more at the start of treatment were reduced to 20% or less after treatment. This finding implies that the threshold for meaningful improvement in SALT scores is 30% for the patient population targeted for participation in the ALLEGRO-2b/3 trial.[8]

However, no studies to date have quantitatively derived thresholds for meaningful hair regrowth using widely accepted and recommended *anchor-based methods*.[9,10] Anchor-based approaches establish a meaningful threshold for within-patient change in an outcome measure by relating that measure to an external anchor; anchors should be understandable and relevant to patients and should characterise the same or a closely related construct.[11] Meaningful within-patient change thresholds can aid in the interpretation of study findings. The objective of this analysis was to estimate a range of potential thresholds indicative of meaningful change in clinician-reported SALT scores using available patient-reported anchor measures representing treatment benefit from the patient perspective.

METHODS

Data source

Data were collected from ALLEGRO-2b/3,[12] an international, randomised, double-blind, placebo-controlled, combined dose-ranging and pivotal study designed to investigate the efficacy of ritlecitinib versus placebo in adults (aged ≥18 years) and adolescents (aged 12–17 years) with a clinical diagnosis of AA and ≥50% hair loss of the scalp (including patients with alopecia totalis or complete hair loss on the scalp; and alopecia universalis, or complete hair loss on the scalp and body). Other key study inclusion criteria were no terminal hair regrowth within 6 months at both the screening and baseline visits and a maximum duration of the current episode of hair loss ≤10 years. Patients were excluded if they had participated in other studies involving investigational drugs within 8 weeks of the ALLEGRO-2b/3 study, had other types of alopecia or other scalp disease that could impact AA assessment, had active systemic diseases that may cause hair loss or had any psychiatric condition.

The maximum duration of participation in ALLEGRO-2b/3 was 57 weeks, including a screening period of up to 5 weeks, a 48-week treatment period and a 4-week follow-up period. The placebo-controlled treatment period was composed of a 4-week loading phase and a 20-week maintenance phase, followed by a 24-week extension phase. Data from patients with SALT scores at baseline, Week 24 and Week 48 were used in the current analysis. Patients from both treatment arms were pooled for the analysis sample.

SALT

The SALT is a clinician-reported quantitative assessment of AA severity based on scalp terminal hair loss at a specific time point (see Table 1). To calculate SALT score, the scalp is first divided into four quadrants: back, top of scalp and both sides. Each of these four quadrants is ascribed a percentage equal to the approximate proportion of scalp surface area covered (24% for back, 40% for top of scalp and 18% for each side). For each quadrant, the percentage hair loss is multiplied by that quadrant's surface area, and the four quadrant scores are summed to yield the SALT score. Scores range from 0 (no scalp hair loss) to 100 (complete scalp hair loss). A patient must have had a SALT score ≥50 at both screening and baseline to be eligible for the study.

Patient-reported measures

To facilitate the estimation of thresholds for meaningful change in SALT scores from the patient perspective, four patient-reported items included in the study were considered as candidate anchors: the Patient Global Impression of Change (PGI-C) item and the three items that comprise the Patient Satisfaction with Hair Growth (P-Sat) questionnaire (Table 1).

PGI-C

For the PGI-C, patients provided an overall retrospective assessment of their AA at postbaseline assessment: 'Since the start of the study, my AA has greatly improved,

TABLE 1 Outcome measures used in the analysis.

Outcome Measure	Items, response scale and scoring
SALT[5] Clinician-reported quantitative assessment of the proportion of scalp hair loss	Percentage hair loss is evaluated in four quadrants, each ascribed a percentage equal to the approximate proportion of scalp surface area covered: ■ Back of scalp, 24% of total scalp area ■ Top of scalp, 40% of total scalp area ■ Left side, 18% of total scalp area For each quadrant, the percentage hair loss is multiplied by that quadrant's surface area, and the four quadrant scores are summed to yield the SALT score (e.g., SALT ≥ 50 is defined as ≥50 scalp without hair) Scores range from 0 (no scalp hair loss) to 100 (complete scalp hair loss)
PGI-C Self-reported, single-item overall retrospective assessment of patient's alopecia areata at postbaseline assessment	Since the start of the study, my alopecia areata has: ■ Greatly improved ■ Moderately improved ■ Slightly improved ■ Not changed ■ Slightly worsened ■ Moderately worsened ■ Greatly worsened ■ Scores range from 1 (greatly improved) to 7 (greatly worsened)
P-Sat Self-reported assessment of patient satisfaction with hair that has regrown since the start of the study in three domains: ■ Amount of hair regrowth ■ Quality of new hair regrowth ■ Overall satisfaction with hair regrowth	How satisfied are you with the *amount* of hair that has grown back since the start of the study? How satisfied are you with the *quality* of the new hair regrowth you have experienced since the start of the study? How *overall* with the hair that has grown back since the start of the study? ■ Very satisfied ■ Moderately satisfied ■ Slightly satisfied ■ Neither satisfied nor dissatisfied ■ Slightly dissatisfied ■ Moderately dissatisfied ■ Very dissatisfied ■ Scores range from 1 (very satisfied) to 7 (very dissatisfied)

Abbreviations: PGI-C, Patient Global Impression of Change; P-Sat, Patient Satisfaction with Hair Growth; SALT, Severity of Alopecia Tool.

moderately improved, slightly improved, not changed, slightly worsened, moderately worsened, greatly worsened'. Scores range from 1 (greatly improved) to 7 (greatly worsened).

P-Sat

The P-Sat asks the patient to evaluate satisfaction with the hair that has regrown since the start of the study. This measure is composed of three items asking about satisfaction with the 'amount' and 'quality' of hair, as well as 'overall' satisfaction with the hair. Scores range from 1 (very satisfied) to 7 (very dissatisfied).

Connection between constructs assessed by the SALT and patient-reported outcome (PRO) measures

Figure 1 graphically diagrams the connection between the constructs assessed by the SALT scores (at two time points, baseline and Week 24) and the constructs assessed by the PRO measures at Week 24. The circles in the figure depict hair growth patterns, on average, at baseline and at Week 24. The SALT change-from-baseline scores shown in the figure represent change in hair loss as well as a change in hair growth (i.e., hair regrowth) at Week 24. Each of the candidate PRO measures addresses treatment benefit from the patient perspective, by asking patients to evaluate either improvement in AA (PGI-C) or satisfaction with hair regrowth (P-Sat items).

Statistical analyses

Analyses were conducted overall and by adult and adolescent subgroups. Missing data were not imputed. To support the use of the PGI-C and each of the three P-Sat items as anchor measures to estimate clinically meaningful change, polyserial correlations between SALT change-from-baseline scores and the four patient-reported items were computed. Correlations that were at least 0.371 in magnitude signified that the proposed anchor measure was acceptable, on the basis of the conversion of the correlation to a large effect on group mean difference (Cohen's $d = 0.80$) for equal group sample sizes, using Cohen's rule of thumb for interpreting effect sizes.[13–15] Descriptive statistics of the SALT change-from-baseline scores were also computed for each anchor measure, to confirm that the size and

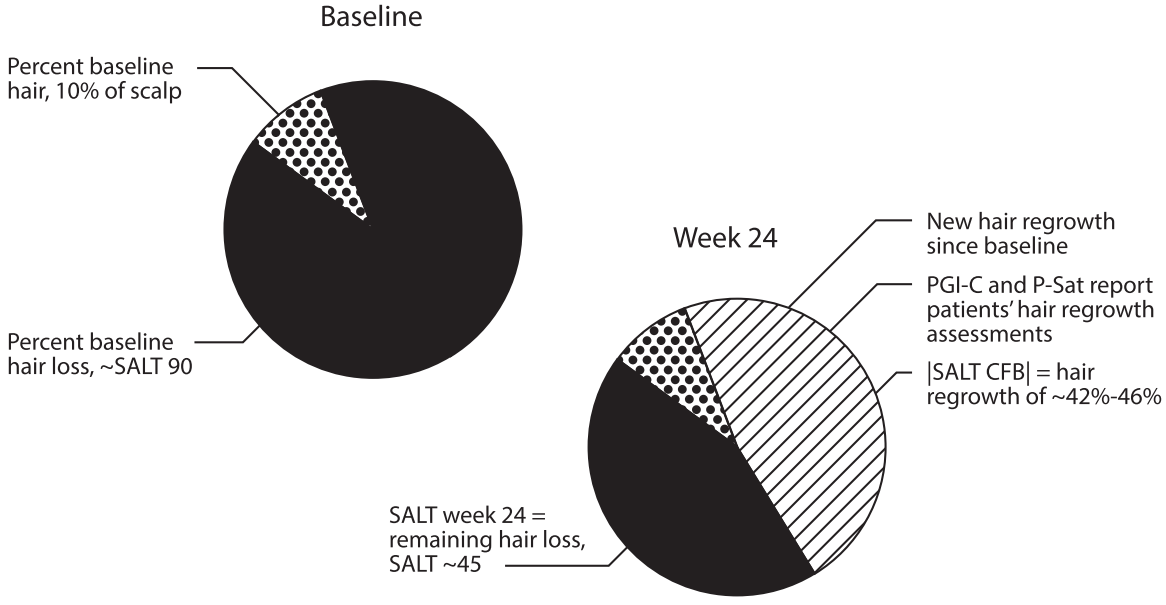

FIGURE 1 Relationship between Severity of Alopecia Tool (SALT) and patient-reported outcome (PRO) measures. The circles in the figure represent a patient's scalp at baseline and at Week 24. At baseline, the dots represent a region of hair covering approximately 10% of the scalp; the black region corresponds to an absolute mean SALT score of 90.6 at baseline. At Week 24, the dotted region continues to represent the baseline condition (10% hair coverage); the black region corresponds to an absolute SALT score of 45; and the lined region represents the amount of hair regrowth. The lined region is equal to the change in hair loss from baseline (SALT 90.6) to Week 24 (SALT 45), yielding a SALT change-from-baseline score of –45. Hence, SALT change-from-baseline scores represent change in hair loss as well as a change in hair growth (i.e., hair regrowth). CFB, change from baseline; PGI-C, Patient Global Impression of Change; P-Sat, Patient Satisfaction with Hair Growth.

direction of the mean and median SALT change-from-baseline scores followed a predictable pattern. For example, greater improvement or worsening in the SALT change-from-baseline scores should be achieved by patients who show greater levels of improvement or worsening on the anchor measures, respectively.

The analysis described in this paper used the PGI-C and P-Sat items to identify four participant subgroups to facilitate threshold estimation: those who chose the 'moderately improved' category on the PGI-C and/or those who chose the 'moderate satisfaction' category on each of the three P-Sat items at Week 24. Moderate improvement as defined by the PGI-C and moderate satisfaction as defined by the P-Sat were selected as benchmarks for patient-reported benefits in this analysis to ensure threshold estimates with the potential to detect benefits that are modest in magnitude yet still meaningful to patients. Considering the constructs addressed by the SALT and the patient-reported items, the PGI-C and P-Sat item pertaining to the amount of hair regrowth were considered primary anchors, whereas the remaining P-Sat items (pertaining to the quality of hair regrowth and overall satisfaction) were considered supportive anchors. Estimates of thresholds for meaningful within-patient change in SALT scores were computed as the mean SALT change-from-baseline scores among patients who reported moderate improvement on the PGI-C or moderate satisfaction on a P-Sat item at week 24.[16,17] These mean scores were reported across all seven PGI-C and P-Sat responses.

TABLE 2 Baseline characteristics among patients with PGI-C or P-Sat ratings and with SALT at Week 24.

Baseline characteristic	
Age	
Mean (SD), *n*	33.6 (14.2), 650
12–17 years, *n* (%)	94 (14.5)
≥18 years, *n* (%)	556 (85.5)
Female, *n* (%)	402 (61.8)
Duration since AA diagnosis, *n* (%)	
<1 year	58 (8.9)
1–<3 years	120 (18.5)
3–7 years	156 (24.0)
>7 years	316 (48.6)
Duration of current AA episode, *n* (%)	
<6 months	42 (6.5)
6–12 months	106 (16.3)
>1–2 years	139 (21.4)
>2 years	363 (55.8)
Baseline SALT scores	
Mean (SD), *n*	90.6 (14.3), 650
50–74, *n* (%)	116 (17.8)
75–99, *n* (%)	237 (36.5)
100, *n* (%)	297 (45.7)

Abbreviations: AA, alopecia areata; PGI-C, Patient Global Impression of Change; P-Sat, Patient Satisfaction with Hair Growth; SALT, Severity of Alopecia Tool; SD, standard deviation.

RESULTS

Baseline characteristics and descriptive SALT scores

Table 2 presents baseline characteristics for the analysis sample of patients with PGI-C or P-Sat ratings and with SALT scores at Week 24. The sample had a mean (standard deviation [SD]) age of 33.6 (14.2) years and was 61.8% female. Roughly half the sample (48.6%) had received their AA diagnosis more than 7 years earlier; for more than half the sample (55.8%), the duration of their current AA episode was longer than 2 years. The 650 participants with SALT scores at week 24 had a baseline mean SALT score of 90.6 (SD, 14.3): 45.7% of participants had a SALT score of 100, or complete scalp hair loss and 17.8% and 36.5% had SALT scores of 50–74 and 75–99, respectively (Table 2). By Week 24, the mean SALT score was 66.4 (SD, 35.6) for the 650 participants with SALT scores. Supporting Information S1: Table S1 presents descriptive SALT scores at baseline and Week 24; Tables 3 and 4, respectively, present absolute SALT scores by PGI-C levels and P-Sat responses at Week 24.

SALT change scores

The 650 participants with SALT scores at week 24 had a mean SALT change-from-baseline score of −24.2 (SD, 31.4). Correlations between SALT change-from-baseline scores between baseline and Week 24 and the PGI-C and P-Sat items at Week 24 ranged from 0.75 to 0.82, well above the 0.371 criterion, providing support for the PGI-C and P-Sat items as anchor measures to empirically establish a threshold for meaningful change in SALT scores (Table 5).

For participants reporting moderate improvement ($n = 102$) on the PGI-C at Week 24, the mean SALT change-from-baseline score was −42.2 (SD, 26.1; 95% confidence interval [CI], −47.3 to −37.1) (Table 6). The pattern of SALT change-from-baseline (to Week 24)

TABLE 3 Absolute SALT scores by PGI-C levels at Week 24.

PGI-C							
SALT score	**Greatly improved**	**Moderately improved**	**Slightly improved**	**Not changed**	**Slightly worsened**	**Moderately worsened**	**Greatly worsened**
Overall							
n	153	102	192	153	20	8	21
Mean (SD)	23.5 (23.4)	45.4 (27.7)	81.9 (22.8)	96.9 (8.6)	85.3 (17.9)	87.9 (13.9)	89.3 (17.9)
Median	14.5	46.6	94.0	100.0	92.6	92.7	99.8
95% CI	19.7–27.2	40.0–50.9	78.7–85.2	95.5–98.2	76.9–93.7	76.3–99.6	81.2–97.5
Missing	1/154 (0.6)	0/102 (0.0)	4/196 (2.0)	1/154 (0.6)	0/20 (0.0)	0/8 (0.0)	0/21 (0.0)
Adult							
n	126	86	163	138	16	8	18
Mean (SD)	22.4 (23.6)	44.9 (27.8)	81.8 (21.9)	96.6 (8.9)	84.8 (19.8)	87.9 (13.9)	87.6 (18.8)
Median	13.9	46.0	91.2	100.0	96.5	92.7	98.3
95% CI	18.2–26.5	39.0–50.9	78.4–85.2	95.1–98.1	74.2–95.3	76.3–99.6	78.2–96.9
Missing	0/126 (0.0)	0/86 (0.0)	1/164 (0.6)	1/139 (0.7)	0/16 (0.0)	0/8 (0.0)	0/18 (0.0)
Adolescent							
n	27	16	29	15	4	0	3
Mean (SD)	28.6 (22.4)	48.2 (27.9)	82.6 (28.0)	99.2 (2.0)	87.4 (8.5)	–	100.0 (0.0)
Median	32.9	50.7	97.1	100.0	83.8	–	100.0
95% CI	19.8–37.5	33.4–63.1	72.0–93.3	98.0–100.3	73.9–100.9	-	-
Missing	1/28 (3.6)	0/16 (0.0)	3/32 (9.4)	0/15 (0.0)	0/4 (0.0)	-	0/3 (0.0)

Note: Green shading highlights mean (SD) SALT score for patients with moderate improvement on the PGI-C.

Abbreviations: CI, confidence interval; PGI-C, Patient Global Impression of Change; SALT, Severity of Alopecia Tool; SD, standard deviation.

TABLE 4 Absolute SALT scores by P-Sat at Week 24: Overall sample.

	P-Sat						
	Very satisfied	**Moderately satisfied**	**Slightly satisfied**	**Neither satisfied or dissatisfied**	**Slightly dissatisfied**	**Moderately dissatisfied**	**Very dissatisfied**
Amount of hair							
n	126	122	102	109	29	39	120
Mean (SD)	22.5 (24.5)	44.8 (29.6)	71.1 (26.1)	91.8 (16.2)	83.8 (25.6)	90.7 (14.3)	94.6 (12.1)
Median	13.5	42.9	76.7	100.0	99.2	99.6	100.0
95% CI	18.2, 26.9	39.4, 50.1	66.0, 76.2	88.7, 94.8	74.1, 93.5	86.0, 95.3	92.4, 96.8
Missing	1/127 (0.8)	1/123 (0.8)	1/103 (1.0)	0/109 (0.0)	1/30 (3.3)	2/41 (4.9)	0/120 (0.0)
Quality of hair							
N	117	123	100	133	29	36	109
Mean (SD)	25.3 (26.0)	41.6 (29.2)	67.0 (29.5)	91.8 (16.1)	80.4 (29.1)	89.7 (15.6)	94.6 (11.9)
Median	14.5	39.0	74.3	100.0	98.0	99.8	100.0
95% CI	20.6, 30.1	36.4, 46.8	61.1, 72.8	89.1, 94.6	69.3, 91.5	84.5, 95.0	92.4, 96.9
Missing	1/118 (0.8)	3/126 (2.4)	0/100 (0.0)	0/133 (0.0)	0/29 (0.0)	2/38 (5.3)	0/109 (0.0)
Overall satisfaction with hair							
n	126	122	103	115	29	34	118
Mean (SD)	23.8 (25.3)	41.9 (28.7)	72.9 (25.3)	92.1 (16.2)	80.7 (27.1)	92.6 (13.5)	94.7 (11.3)
Median	14.3	38.8	82.1	100.0	99.0	100.0	100.0
95% CI	19.4, 28.3	36.7, 47.0	68.0, 77.9	89.1, 95.1	70.4, 91.0	87.9, 97.3	92.6, 96.7
Missing	2/128 (1.6)	1/123 (0.8)	0/103 (0.0)	1/116 (0.9)	0/29 (0.0)	2/36 (5.6)	0/118 (0.0)

Note: Green shading highlights mean (SD) SALT score for patients with moderate satisfaction on the P-Sat. Supporting Information S1: Tables S2 and S3, respectively, present absolute SALT scores by P-Sat at Week 24 for adults and adolescents.

Abbreviations: CI, confidence interval; P-Sat, Patient Satisfaction with Hair Growth; SALT, Severity of Alopecia Tool; SD, standard deviation.

TABLE 5 Correlations between SALT scores and PGI-C and P-Sat items at Week 24.

	Correlation coefficient with SALT			
	PGI-C	**P-SAT amount of scalp hair**	**P-SAT quality of scalp hair**	**P-SAT overall satisfaction with scalp hair**
Correlations with SALT score				
Overall				
Polyserial correlation (95% CI), *n*	0.78* (0.74–0.81), 649	0.76* (0.72–0.79), 647	0.73* (0.69–0.77), 647	0.76* (0.72–0.79), 647
Interpretation	Strong	Strong	Strong	Strong
Adults				
Polyserial correlation (95% CI), *n*	0.78* (0.74, 0.82), 555	0.76* (0.72, 0.80), 554	0.73* (0.69, 0.77), 554	0.76* (0.72, 0.80), 554
Interpretation	Strong	Strong	Strong	Strong
Adolescents				
Polyserial correlation (95% CI), *n*	0.77* (0.68, 0.86), 94	0.78* (0.69, 0.86), 93	0.75* (0.65, 0.84), 93	0.76* (0.67, 0.85), 93
Interpretation	Strong	Strong	Strong	Strong
Correlations with change in SALT score				
Overall				
Polyserial correlation (95% CI), *n*	0.82* (0.79, 0.85), 649	0.78* (0.74, 0.81), 647	0.75* (0.71, 0.78), 647	0.78* (0.75, 0.81), 647
Interpretation	Strong	Strong	Strong	Strong
Adults				
Polyserial correlation (95% CI), *n*	0.83* (0.80, 0.86), 555	0.78* (0.74, 0.82), 554	0.75* (0.71, 0.79), 554	0.78* (0.75, 0.82), 554
Interpretation	Strong	Strong	Strong	Strong
Adolescents				
Polyserial correlation (95% CI), *n*	0.79* (0.71, 0.88), 94	0.78* (0.70, 0.87), 93	0.76* (0.66, 0.85), 93	0.79* (0.71, 0.88), 93
Interpretation	Strong	Strong	Strong	Strong

Abbreviations: CI, confidence interval; PGI-C, Patient Global Impression of Change; P-Sat, Patient Satisfaction with Hair Growth; SALT, Severity of Alopecia Tool.

*$p < 0.05$.

TABLE 6 Change from baseline in SALT scores by PGI-C levels at Week 24.

PGI-C							
Change in SALT score	**Greatly improved**	**Moderately improved**	**Slightly improved**	**Not changed**	**Slightly worsened**	**Moderately worsened**	**Greatly worsened**
Overall							
n	153	102	192	153	20	8	21
Mean (SD)	−62.4 (23.1)	−42.2 (26.1)	−11.1 (16.5)	−0.6 (4.1)	2.4 (9.4)	9.1 (12.4)	10.0 (22.0)
Median	−62.0	-43.8	−3.3	0.0	0.2	8.5	10.6
95% CI	−66.1, −58.7	−47.3, −37.1	−13.4, −8.7	−1.2, 0.1	−1.9, 6.8	−1.3, 19.4	0.0, 20.0
Missing	1/154 (0.6)	0/102 (0.0)	4/196 (2.0)	1/154 (0.6)	0/20 (0.0)	0/8 (0.0)	0/21 (0.0)
Adult							
n	126	86	163	138	16	8	18
Mean (SD)	−62.9 (22.8)	−42.9 (25.3)	−11.1 (16.0)	−0.6 (4.2)	0.5 (8.4)	9.1 (12.4)	10.1 (23.4)
Median	−62.4	−44.1	−3.4	0.0	0.1	8.5	11.8
95% CI	−66.9, -58.9	−48.3, −37.4	−13.6, −8.7	−1.3, 0.1	−4.0, 4.9	−1.3, 19.4	−1.5, 21.8
Missing	0/126 (0.0)	0/86 (0.0)	1/164 (0.6)	1/139 (0.7)	0/16 (0.0)	0/8 (0.0)	0/18 (0.0)
Adolescent							
n	27	16	29	15	4	0	3
Mean (SD)	−60.0 (24.6)	−38.5 (30.4)	−10.6 (19.4)	0.1 (3.1)	10.3 (10.1)	-	9.0 (12.8)
Median	−57.3	−36.6	−2.9	0.0	8.6	-	3.3
95% CI	−69.7, −50.3	−54.8, −22.3	−18.0, −3.2	−1.6, 1.8	−5.8, 26.4	-	−22.9, 40.8
Missing	1/28 (3.6)	0/16 (0.0)	3/32 (9.4)	0/15 (0.0)	0/4 (0.0)	-	0/3 (0.0)

Note: Green shading highlights mean (SD) change from baseline SALT score for patients with moderate improvement on the PGI-C.

Abbreviations: CI, confidence interval; PGI-C, Patient Global Impression of Change; SALT, Severity of Alopecia Tool; SD, standard deviation.

scores was consistent across improved levels of PGI-C for both adults and adolescents; however, the mean SALT change-from-baseline score for adults reporting moderate improvement ($n = 86$) was slightly larger (−42.9 [SD, 25.3; 95% CI, −48.3 to −37.4]) than for adolescents ($n = 16$) reporting moderate improvement (−38.5 [SD, 30.4; 95% CI, −54.8 to −22.3]) (Table 6).

For participants reporting moderate satisfaction with the amount of hair growth ($n = 122$; P-Sat item 1), the quality of hair growth ($n = 123$; P-Sat item 2) or overall hair growth ($n = 122$; P-Sat item 3), mean SALT change-from-baseline scores were −43.1 (SD, 26.8; 95% CI, −47.9 to −38.3), −45.9 (SD, 27.1; 95% CI, −50.7 to -41.0) and -45.2 (SD, 25.9; 95% CI, −49.8 to −40.5), respectively (Table 7). Among participants in the very satisfied to neither satisfied or dissatisfied range, the pattern of SALT change-from-baseline scores was consistent across levels of the P-Sat items in the adult and adolescent groups with adequate sample sizes (Supporting Information S1: Tables S4 and S5).

Based on the rounded mean SALT change-from-baseline score for patients (overall, including both adults and adolescents) reporting moderate improvement in AA or moderate satisfaction in the amount of hair regrowth, the primary anchors for these analyses, the thresholds for meaningful change in SALT scores were 42 and 43, respectively. For patients reporting moderate satisfaction in the quality of hair regrowth and regrowth overall at Week 24, the thresholds defining meaningful change in hair regrowth were SALT 46 and 45, respectively.

DISCUSSION

This analysis demonstrates that changes in clinicians' assessments of hair loss were strongly associated with patients' perceptions of improvement in AA and satisfaction with three aspects of hair regrowth (amount, quality and overall) following treatment. These associations support the use of the patient-reported PGI-C and P-Sat items as anchors to identify thresholds for meaningful improvements in clinician-reported SALT scores. Based on the mean SALT change-from-baseline scores for patients with AA reporting moderate improvement in AA or moderate satisfaction in aspects of hair regrowth (amount, quality, overall) at Week 24, threshold estimates for defining meaningful levels of improvement in SALT scores ranged from 42 to 46. Results for adult and adolescent subgroups were consistent with the overall findings.

While qualitative research with a real-world sample has suggested that a 30% reduction among patients with baseline SALT scores of 50 or greater would represent meaningful improvement,[8] this evidence relied on asking patients about hypothetical levels of potential but unrealised hair regrowth. To our knowledge, the present analysis is the first to empirically estimate thresholds for meaningful clinical response in AA, as indicated by SALT change-from-baseline scores, from patients' real experiences of hair growth after receiving therapy for extensive AA. Importantly, the analyses here reflect the experiences of patients who reported moderate improvement in AA after 24 weeks of treatment; thus, improvements of 42–46 in SALT scores represent meaningful thresholds for patients who have started treatment. These patients may anticipate further improvements in SALT scores as treatment continues, with the aim of achieving of SALT scores of 20 or below, the primary endpoint used by the FDA for the ALLEGRO-2b/3 trial. A SALT score of 20 or below clearly reflects clinically meaningful change, and our analyses support this endpoint as an ultimate goal of treatment.

As noted above, the average SALT score at baseline within the analysis sample was 90.6 (SD, 14.3), and 45.7% of patients had complete scalp hair loss. On average, participants in the trial would then need to improve by their SALT score by approximately 70 to meet the trial endpoint (i.e., achieve a score of 20). Similarly, 82.2% of participants had SALT scores of at least 75 at baseline and would require a reduction of 55 in their SALT scores to be classified as treatment responders. Of note, at Week 24, patients who reported being very satisfied with amount of hair growth, quality of hair growth, or overall hair growth on the P-Sat had absolute mean SALT scores ranging from 22.5 to 25.3, reflecting that achievement of a SALT score close to 20 represents a substantial treatment benefit from the patient perspective.

This analysis has a number of strengths, most notably the use of longitudinal data for standardised endpoints collected at regular intervals in a trial setting where patients received active therapy to promote hair regrowth. Response rates on the patient-reported measures were high, and rates of missing data were low. The analysis also statistically confirmed the appropriateness of anchors. Nonetheless, limitations are acknowledged. One limitation is that the of available measures for anchors (PGI-C and P-Sat) in ALLEGRO-2b/3 were developed specifically to assess patient perspective on their change in AA and treatment satisfaction. Anchors with underlying constructs more closely targeted to hair loss or hair regrowth (as opposed to satisfaction with regrowth) could strengthen the results, although such anchors are, to our knowledge, lacking. Another potential issue is the use of a retrospective global measure (PGI-C) as an anchor rather than an anchor measuring the change in global status of severity.[9] Finally, these

TABLE 7 Change from baseline in SALT scores by P-Sat at Week 24: Overall sample.

P-Sat							
Change in SALT score	**Very satisfied**	**Moderately satisfied**	**Slightly satisfied**	**Neither satisfied or dissatisfied**	**Slightly dissatisfied**	**Moderately dissatisfied**	**Very dissatisfied**
Amount of hair							
n	126	122	102	109	29	39	120
Mean (SD)	−63.5 (23.7)	−43.1 (26.8)	−18.7 (20.3)	−2.7 (10.6)	−11.1 (21.2)	−2.4 (11.1)	1.4 (11.5)
Median	−66.3	−44.1	−10.4	0.0	0.0	0.0	0.0
95% CI	−67.7, −59.3	−47.9, −38.3	−22.7, −14.7	−4.7, −0.7	−19.1, −3.0	−6.1, 1.2	−0.7, 3.5
Missing	1/127 (0.8)	1/123 (0.8)	1/103 (1.0)	0/109 (0.0)	1/30 (3.3)	2/41 (4.9)	0/120 (0.0)
Quality of hair							
N	117	123	100	133	29	36	109
Mean (SD)	−60.7 (25.1)	−45.9 (27.1)	−23.6 (24.3)	−2.9 (10.6)	−11.5 (21.7)	−2.5 (11.6)	1.8 (11.5)
Median	−62.4	−47.2	−14.4	0.0	−1.0	0.0	0.0
95% CI	−65.3, −56.1	−50.7, −41.0	−28.4, −18.8	−4.7, −1.1	−19.7, −3.2	−6.4, 1.4	−0.4, 4.0
Missing	1/118 (0.8)	3/126 (2.4)	0/100 (0.0)	0/133 (0.0)	0/29 (0.0)	2/38 (5.3)	0/109 (0.0)
Overall satisfaction with hair							
N	126	122	103	115	29	34	118
Mean (SD)	−62.6 (24.9)	−45.2 (25.9)	−18.3 (19.6)	−2.0 (10.2)	−12.7 (20.8)	−2.8 (10.8)	2.1 (10.3)
Median	−66.3	−48.4	−11.0	0.0	−0.2	0.0	0.0
95% CI	−66.9, −58.2	−49.8, −40.5	−22.1, −14.5	−3.9, -0.1	−20.6, −4.8	−6.6, 1.0	0.3, 4.0
Missing	2/128 (1.6)	1/123 (0.8)	0/103 (0.0)	1/116 (0.9)	0/29 (0.0)	2/36 (5.6)	0/118 (0.0)

Note: Green shading highlights mean (SD) change from baseline SALT score for patients with moderate satisfaction on the P-Sat. Tables S4 and S5, respectively, present change from baseline SALT scores by P-Sat at week 24 for adults and adolescents.

Abbreviations: CI, confidence interval; P-Sat, Patient Satisfaction with Hair Growth; SALT, Severity of Alopecia Tool; SD, standard deviation.

analyses focus on ritlecitinib, limiting generalisability to other AA treatments. Future analyses that include a broader range of treatments or comparisons between active treatments could yield a more comprehensive understanding of treatment benefits across interventions.

In conclusion, threshold estimates signifying meaningful improvements in SALT scores were consistent across anchors representing the patient perspective, ranging from absolute SALT scores of 42–46 on the basis of analysis of the ALLEGRO-2b/3 trial data. While the achievement of low SALT scores of ≤10 and ≤20 have been used to characterise treatment benefit in clinical trials, the amount of change required to meet this endpoint far exceeds the estimates computed in our study, as well as those suggested by qualitative research.[8] The results of this study serve as a useful reminder that the treatment goals of individual patients must be considered when evaluating benefit in clinical practice.

AUTHOR CONTRIBUTIONS

Ernest H. Law, Nicole J. Williams, Dane Korver, Randall H. Bender, Debanjali Mitra, Gregor Schaefer and Lauren M. Nelson each made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted and agree to be accountable for all aspects of the work.

ACKNOWLEDGEMENTS

Kate Lothman of RTI Health Solutions provided medical writing services for the development of this manuscript. These services were funded by Pfizer Inc.

CONFLICTS OF INTEREST STATEMENT

This study was conducted under a research contract between RTI Health Solutions and Pfizer Inc. and was funded by Pfizer Inc. Ernest H. Law, Debanjali Mitra and Gregor Schaefer are employees of Pfizer Inc., the sponsor of this study. Dane Korver, Randall H. Bender and Lauren M. Nelson are employees of RTI Health Solutions. Nicole J. Williams was an employee of RTI Health Solutions when this research was conducted.

DATA AVAILABILITY STATEMENT

Upon request, and subject to review, Pfizer will provide the data that support the findings of this study. Subject to certain criteria, conditions and exceptions, Pfizer may also provide access to the related individual deidentified participant data. See https://www.pfizer.com/science/clinical-trials/trial-data-and-results for more information.

ETHICS STATEMENT

The study underlying this analysis, NCT03732807, was approved by the institutional review boards or ethics committees of the participating institutions. The study was conducted in accordance with the general principles set forth in the International Ethical Guidelines for Biomedical Research Involving Human Subjects (Council for International Organizations of Medical Sciences 2002), ICH Guideline for Good Clinical Practice and the Declaration of Helsinki. Written informed consent was obtained from each patient, parent, or the patient's legal representative.

ORCID

Ernest H. Law http://orcid.org/0000-0002-6111-8008

REFERENCES

1. Perera E, Yip L, Sinclair R. Alopecia areata. Curr Probl Dermatol. 2015;47:67–75.
2. Cash TF. The psychological effects of androgenetic alopecia in men. J Am Acad Dermatol. 1992;26(6):926–31.
3. Cash TF, Price VH, Savin RC. Psychological effects of androgenetic alopecia on women: comparisons with balding men and with female control subjects. J Am Acad Dermatol. 1993;29(4):568–75.
4. Ruiz-Doblado S, Carrizosa A, García-Hernández MJ. Alopecia areata: psychiatric comorbidity and adjustment to illness. Int J Dermatol. 2003;42(6):434–7.
5. Olsen E, Hordinsky M, McDonald-Hull S, Price V, Roberts J, Shapiro J, et al. Alopecia areata investigational assessment guidelines. J Am Acad Dermatol. 1999;40(2 Pt 1):242–6.
6. Olsen EA, Hordinsky MK, Price VH, Roberts JL, Shapiro J, Canfield D, et al. Alopecia areata investigational assessment guidelines–part II. J Am Acad Dermatol. 2004;51(3):440–7.
7. Olsen EA, Roberts J, Sperling L, Tosti A, Shapiro J, McMichael A, et al. Objective outcome measures: collecting meaningful data on alopecia areata. J Am Acad Dermatol. 2018;79(3):470–8.e3.
8. Wyrwich KW, Kitchen H, Knight S, Aldhouse NVJ, Macey J, Nunes FP, et al. The Alopecia Areata Investigator Global Assessment scale: a measure for evaluating clinically meaningful success in clinical trials. Br J Dermatol. 2020;183(4):702–9.
9. McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. Expert Rev Pharmacoecon Outcomes Res. 2011;11(2):163–9.
10. McLeod LD, Cappelleri JC, Hays RD. Best (but oft-forgotten) practices: expressing and interpreting associations and effect sizes in clinical outcome assessments. Am J Clin Nutr. 2016;103(3):685–93.
11. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. Mayo Clin Proc. 2002;77(4):371–83.
12. ClinicalTrials.gov. PF-06651600 for the treatment of alopecia areata: ALLEGRO-2b/3. https://clinicaltrials.gov/ct2/show/NCT03732807 (2021). Accessed 9 Feb 2022.

13. Hays R, Revicki DA. Reliability and validity (including responsiveness). In: Fayers P, Hays R, editors. Assessing quality of life in clinical trials. New York: Oxford University Press; 2005. p. 25–39.
14. Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. J Chron Obstruct Pulmon Dis. 2005;2(1): 63–7. https://doi.org/10.1081/copd-200050663
15. Fayers PM, Hays RD. Don't middle your MIDs: regression to the mean shrinks estimates of minimally important differences. Qual Life Res. 2014;23(1):1–4. https://doi.org/10.1007/s11136-013-0443-4
16. Food and Drug Administration (FDA). Patient-focused drug development (PFDD) guidance: methods to identify what is important to patients & select, develop or modify fit-for-purpose clinical outcomes assessments. Attachment to guidance 3 discussion document—appendices. https://www.fda.gov/drugs/newsevents/ucm607276.htm (2018). Accessed 11 Aug 2021.
17. Food and Drug Administration (FDA). Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf (2009). Accessed 11 Aug 2021.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Law EH, Williams NJ, Korver D, Bender RH, Mitra D, Schaefer G, et al. Determining meaningful thresholds for evaluating treatment efficacy in patients with alopecia areata. JEADV Clin Pract. 2024;3:1119–31. https://doi.org/10.1002/jvc2.441